

PLANT STEROL REDUCTASES AND USES THEREOF

5

Background of the Invention

This application relates to plant sterol biosynthetic enzymes, genes, and their uses.

Plant sterols belong to a large group of secondary compounds known as
10 terpenes or isoprenoids. Sterol biosynthesis in plants generally involves a series of different enzymatic steps in the isoprenoid pathway that result in the formation of a variety of sterol end products (Benveniste *Ann. Rev. Biochem.* 37:275, 1986). Although such sterol compounds have been identified in higher plants, their function in plant growth and development is poorly understood.

15 One such plant sterol, brassinolide, that belongs to a class of sterols referred to as brassinosteroids (BR), was first discovered in the pollen of *Brassica napus* (Grove et. al., *Nature* 281: 216, 1979). Brassinosteroids are growth-promoting natural products having structural similarities to animal steroid hormones. The wide distribution of brassinosteroids in the plant kingdom, their effect on cell proliferation
20 and elongation, and their interactions with other plant hormones (e.g., cytokinins), have indicated that these compounds are plant-growth regulators. Brassinosteroids are thought to promote hypocotyl elongation, leaf unrolling, and xylem differentiation. In addition, such compounds are also believed to be involved in de-etiolation of cotyledons, root elongation, radial growth, and anthocyanin formation.

25 The function of plant sterol growth regulators, such as BR, in relationship to other classes of plant growth regulators such as auxin, gibberellin, abscisic acid, and cytokinin, during plant development also needs to be evaluated. For example, the growth regulator, cytokinin, is known to affect a variety of developmental processes including photomorphogenesis, chloroplast biogenesis and maintenance, apical

dominance, and senescence. In addition, this growth regulator is thought to antagonize BR's ability to promote hypocotyl elongation and cotyledon de-etiolation.

Summary of the Invention

5 In general, the invention features a substantially pure plant C-14 sterol reductase polypeptide. Preferably, the C-14 sterol reductase polypeptide includes an amino acid sequence substantially identical to the sequence shown in Fig. 14 (SEQ ID NO: 1); and is from a dicot (for example, a crucifer or a solanaceous plant), monocot, gymnosperm, or an alga.

10 In related aspects, the invention features purified DNA that includes a sequence encoding a C-14 sterol reductase polypeptide (for example, a sequence substantially identical to the DNA sequence shown in Fig. 14; SEQ ID NO: 2; or a DNA sequence that encodes a C-14 sterol reductase polypeptide which has an amino acid sequence substantially identical to that shown in Fig. 14; SEQ ID NO: 1). The invention also
15 features a vector and a cell, each of which includes purified DNA encoding a C-14 sterol reductase polypeptide; and a method of producing a recombinant C-14 sterol reductase polypeptide involving providing a cell (for example, a plant cell) transformed with purified DNA encoding a C-14 sterol reductase polypeptide positioned for expression in the cell, culturing the transformed cell under conditions
20 for expressing the DNA, and isolating the recombinant C-14 sterol reductase polypeptide. The invention further features recombinant C-14 sterol reductase produced by such expression of a purified DNA, and an isolated antibody that specifically recognizes and binds a plant C-14 sterol reductase polypeptide.

25 In addition, the invention features nucleotide sequences that hybridize to a C-14 sterol reductase gene (including the coding sequence of such a gene and its complement) and that encode a C-14 sterol reductase polypeptide. Furthermore, the invention includes oligonucleotide probes that detect a C-14 sterol reductase gene or functional equivalents thereof in a plant (for example, dicots (such as solanaceous and

cruciferous plants), monocots, gymnosperms, and algae). Such probes are useful to isolate DNA sequences that encode C-14 sterol reductases from other plants. In one particular example, oligonucleotides may be designed based on a C-14 sterol reductase sequence disclosed herein and used as hybridization probes or as primers in
5 polymerase chain reactions (PCR). Conserved regions in the C-14 sterol reductase gene are useful in the design of such primers to facilitate the recovery of C-14 sterol reductases from other related and unrelated plants.

In yet other related aspects, the invention features a transgenic plant (or seeds or cells thereof) containing DNA encoding a C-14 sterol reductase polypeptide
10 integrated into the genome of the plant, where the DNA is expressed in the transgenic plant, resulting in the production of a C-14 sterol reductase polypeptide.

In still another aspect, the invention features a method for reducing the level of a plant C-14 sterol reductase polypeptide in a transgenic plant cell. This method generally involves expressing in the transgenic plant cell an antisense C-14 sterol
15 reductase polypeptide nucleic acid sequence. In general, such an antisense C-14 sterol reductase nucleic acid sequence is encoded by a transgene integrated into the genome of the transgenic plant cell and is based on the nucleotide sequence that is shown in Fig. 14 (SEQ ID NO: 2) or Fig. 15. (SEQ ID NO: 3). In preferred embodiments, the plant cell expressing an antisense C-14 sterol reductase nucleic acid sequence is a
20 dicot (for example, crucifer), monocot, gymnosperm, or algal cell. In yet other preferred embodiments, the method involves growing a transgenic plant from the transgenic plant cell, whereby the level of the C-14 sterol reductase polypeptide is reduced in the transgenic plant.

In other related aspects, the invention features a plant cell expressing an
25 antisense C-14 sterol reductase nucleic acid sequence and a plant expression vector that includes an antisense C-14 sterol reductase nucleic acid sequence, where the antisense sequence is operably linked to an expression control region.

- 4 -

In another aspect, the invention features a method for increasing the level of a C-14 sterol reductase in a transgenic plant cell. This method involves expressing in the transgenic plant cell a C-14 sterol reductase polypeptide nucleic acid sequence. Preferably, the method utilizes a C-14 sterol reductase nucleic acid sequence that is substantially identical to the nucleotide sequence that is shown Fig. 14 (SEQ ID NO: 2). In preferred embodiments, the plant cell expressing a C-14 sterol reductase polypeptide nucleic acid sequence is a dicot (for example, a crucifer), monocot, gymnosperm, or algal cell.

In another aspect, the invention features a transgenic plant having a knockout mutation in DNA encoding a plant C-14 sterol reductase polypeptide. Such knockout genes are constructed according to conventional methods (e.g., Lee et al. *Plant Cell* 2: 415, 1990; Miao and Lam, *Plant J.* 7: 359, 1995).

By "plant C-14 sterol reductase" is meant an amino acid sequence that catalyzes the reduction of any sterol precursor having a C14=C15 double bond, for example, as described by Benveniste, *Annu. Rev. Biochem.* 37: 275, 1986. Preferably, such a polypeptide has an amino acid sequence which is at least 30%, preferably 40%, and most preferably 50% or even 80-95% identical to the amino acid sequence of the C-14 sterol reductase polypeptide shown in Fig. 14 (SEQ ID NO: 1). The length of comparison of amino acid sequences will generally be at least 16 amino acids, preferably at least 20 amino acids, more preferably at least 25 amino acids, and most preferably at least 35 amino acids.

By "polypeptide" or "protein" is meant any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation).

By a "substantially identical" polypeptide sequence is meant an amino acid sequence that differs only by conservative amino acid substitutions, for example, substitution of one amino acid for another of the same class (e.g., valine for glycine, arginine for lysine, etc.) or by one or more non-conservative substitutions, deletions,

or insertions, located at positions of the amino acid sequence that do not destroy the function of the polypeptide (assayed, for example, as described herein).

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group
5 (University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705), BLAST, or PILEUP/PRETTYBOX programs). Such software matches sequences by assigning degrees of homology to various substitutions, deletions, and other modifications.

By "substantially pure polypeptide" is meant a polypeptide preparation that is at
10 least 60% by weight (dry weight) the compound of interest, for example, the C-14 sterol reductase polypeptide or C-14 sterol reductase-specific antibody. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight the compound of interest. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel
15 electrophoresis, or HPLC analysis.

By "purified DNA" is meant DNA that is not immediately contiguous with both of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally-occurring genome of the organism from which it is derived. The term therefore includes, for example, a recombinant DNA that is
20 incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote, or that exists as a separate molecule (for example, a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding one or more additional amino
25 acids.

By a "substantially identical" nucleic acid is meant a nucleic acid sequence that encodes a polypeptide differing only by conservative amino acid substitutions, for example, substitution of one amino acid for another of the same class (e.g., valine for

- 6 -

glycine, arginine for lysine, etc.) or by one or more non-conservative substitutions, deletions, or insertions, located at positions of the amino acid sequence that do not destroy the function of the polypeptide (assayed, for example, as described herein). Again, the encoded sequence is at least 30%, more preferably 40%, and most preferably 50%, or even 80 to 95% identical at the amino acid level to the sequence of Fig. 14 (SEQ ID NO: 1). Thus, when nucleic acid sequences are compared, a "substantially identical" nucleic acid sequence is one which is at least 30%, more preferably 40%, and most preferably 50%, or even 80 to 95% identical to the sequence of Fig. 14 (SEQ ID NO: 2). The length of nucleic acid sequence comparison will generally be at least 30 nucleotides, preferably at least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably 110 nucleotides. Again, identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705).

By "isolated antibody" is meant antibody that is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, antibody.

By "specifically binds" is meant an antibody that recognizes and binds a C-14 sterol reductase polypeptide but which does not substantially recognize and bind other molecules in a sample (e.g., a biological sample) which naturally includes a C-14 sterol reductase. An antibody which "specifically binds" a C-14 sterol reductase is sufficient to detect a C-14 sterol reductase product in such a biological sample using one or more of the standard immunological techniques available to those in the art (for example, Western blotting or immunoprecipitation).

By "an antisense C-14 sterol reductase sequence" is meant a nucleotide sequence that is complementary to a plant C-14 sterol reductase messenger RNA. In general, such an antisense sequence will usually be at least 15 nucleotides, preferably

about 15-200 nucleotides, and more preferably 200-2,000 nucleotides in length. The antisense sequence may be complementary to all or a portion of the plant C-14 sterol reductase mRNA nucleotide sequence, and, as appreciated by those skilled in the art, the particular site or sites to which the antisense sequence binds as well as the length of the antisense sequence will vary, depending upon the degree of inhibition desired and the uniqueness of the antisense sequence. By binding to the appropriate target sequence, an RNA-RNA, DNA-DNA, or RNA-DNA duplex is formed. A transcriptional construct expressing a plant C-14 sterol reductase antisense nucleotide sequence includes, in the direction of transcription, a promoter, the sequence coding for the antisense RNA on the sense strand, and a transcriptional termination region. Antisense C-14 sterol reductase sequences may be constructed and expressed as described herein or as described, for example, in van der Krol et al., *Gene* 72: 45, 1988; Rodermeel et al., *Cell* 55: 673, 1988; Mol et al., *FEBS Lett.* 268: 427, 1990; Weigel and Nilsson, *Nature* 377: 495, 1995; Cheung et al., *Cell* 82, 383, 1995; and U.S. Pat. No. 5,107,065. In addition, C-14 sterol reductase antisense sequences are useful for the formation of triple helices, where the antisense sequence is bound to a DNA duplex. By binding to the target nucleic acid, C-14 sterol reductase antisense sequences can inhibit the function of the target nucleic acid. This results, for example, in the blocking of transcription, processing of poly A⁺ addition, replication, translation, or promoting inhibitory mechanisms of the cell, such as RNA degradation. The triple helix-forming and antisense C-14 sterol reductase sequences are useful for selectively suppressing certain cellular functions that are associated with C-14 sterol reductase activity.

By a "transformed cell" is meant a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule encoding (as used herein) a C-14 sterol reductase polypeptide (for example, a substantially identical DNA encoding the C-14 sterol reductase shown in Fig. 14 (SEQ ID NO: 2)).

By "positioned for expression" is meant that the DNA molecule is positioned adjacent to a DNA sequence which directs transcription and translation of the sequence (for example, facilitates the production of, for example, a plant C-14 sterol reductase polypeptide such as the amino acid sequence shown in Fig. 14 (SEQ ID NO: 1)), or an RNA molecule (for example, an antisense RNA).

By "promoter" is meant a minimal sequence sufficient to direct transcription. Included in the invention are promoter elements that are sufficient to render promoter-dependent gene expression controllable for cell-, tissue-, or organ-specific gene expression, or elements that are inducible by external signals or agents (for example, light-, pathogen-, wound-, stress-, or hormone-inducible elements); such elements may be located in the 5' or 3' regions of the native gene or engineered into a transgene construct.

By "operably linked" is meant that a gene and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (for example, transcriptional activator proteins) are bound to the regulatory sequence(s).

By "crucifer" is meant any plant that is classified within the Cruciferae family as commonly described in, e.g., Gray's Manual of Botany American Book Company, N.Y., 1950; *Hortus Third: A Concise Dictionary of Plants Cultivated in the U.S. and Canada*, Macmillan, 1976; or Simmons, N.W., *Evolution of Crop Plants*, 1986. The Cruciferae include many agricultural crops, including, but not limited to, broccoli, cabbage, brussel sprouts, rapeseed, kale, Chinese kale, cauliflower, horseradish, and *Arabidopsis*.

By "plant cell" is meant any self-propagating cell bounded by a semi-permeable membrane and containing a plastid. Such a cell also requires a cell wall if further propagation is desired. Plant cell, as used herein includes, without limitation, algae, cyanobacteria, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores.

- 9 -

By "transgene" is meant any piece of DNA that is inserted by artifice into a cell, and becomes part of the genome of the organism which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (i.e., foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism.

By "transgenic" is meant any cell that includes a DNA sequence which is inserted by artifice into a cell and becomes part of the genome of the organism which develops from that cell. As used herein, the transgenic organisms are generally transgenic plants and the DNA (transgene) is inserted by artifice into the nuclear or plastidic genomes.

Other features and advantages of the invention will be apparent from the following detailed description thereof, and from the claims.

Detailed Description

The drawings will first be described.

Drawings

Figs. 1A-1B are photographs showing that the *ell* mutant morphology was phenocopied by treating wild-type seedlings with 30 μ M of dimethylallylamino-purine (2ip), a synthetic cytokinin, in the dark (**Fig. 1A**) or light (**Fig. 1B**). From left to right in both **Figs. 1A-1B**: wild-type plant; wild-type plant + 2ip; *ell*; and *ell* + 2ip.

Fig. 2 is a photograph illustrating the constitutive photomorphogenesis of *ell* seedling development in the dark. Wild-type (left) and *ell* (right) seedlings were grown in the dark for twenty-one days on Murashige-Skoog (MS) plates containing two percent sucrose.

Figs. 3A-3B are photographs showing that the rosette leaves of the *ell* plant (**Fig. 3B**) are darker green in color than those of the wild-type plant (**Fig. 3A**).

Fig. 4 is a photograph illustrating that an *ell* mutant has reduced apical dominance in comparison to a wild-type plant. Six-week-old wild-type (left) and ten-week-old *ell* (right) plants were grown in the greenhouse.

Fig. 5 is a photograph showing that *ell* mutants (right) exhibit irregular, thickened cotyledons and hypocotyls, and reduced cotyledon petioles compared to wild-type plants (left).

Figs. 6A-6B are photographs showing abnormal flower development in the *ell* mutant. **Fig. 6A** shows, from left to right, that the sepal, petal, stamen, and carpel are shorter in *ell* (lower row) than wild-type (upper row) plants. **Fig. 6B** shows, from left to right, the top and side view of wild-type (left) and *ell* (right) flowers.

Figs. 7A-7F are photographs showing embryo development in *ell* and wild-type plants. Wild-type and *ell* plants are shown in the left and right of each photograph, respectively. **Fig. 7A** shows *ell* embryo development at the 32- to 64-cell stage, and **Fig. 7B** shows that, when wild-type embryos have reached the heart stage, *ell* embryos are only at the globular stage. As shown in **Fig. 7C** and **Fig. 7D**, when the wild-type embryo reached the torpedo stage, the *ell* mutant embryo was at the heart stage. **Fig. 7E** shows that apical hooks were not formed in *ell* embryos. And

Fig. 7F shows that *ell* seeds desiccated without completing the late stages of embryogenesis.

Figs. 8A-8B are photographs of dry seeds from wild-type (Fig. 8A) and *ell* plants (Fig. 8B). Reduced seed size, wrinkled seed coat, and precocious germination were observed in *ell* seeds.

Figs. 9A-9F are photographs showing the supernumerary cotyledons that were observed in the *ell* mutant, including one (Fig. 9A), two (Fig. 9B), three (Fig. 9C), four (Fig. 9D), five (Fig. 9E), and more than six cotyledons (Fig. 9F).

Figs. 10A-10B are illustrations showing various aspects of the molecular characterization of the *Arabidopsis thaliana* C-14 sterol reductase gene. Fig. 10A is a schematic illustration showing the position of a T-DNA insertion into chromosome 3 of *Arabidopsis*, approximately forty base pairs upstream of the *ELL* gene, and the exon-intron structure of the C-14 sterol reductase gene.

Fig. 10B is a schematic illustration showing the map position of *ELL* on chromosome 3.

Fig. 11 is a schematic illustration showing a comparison of the predicated *ELL* amino acid sequence (designated Ath; SEQ ID NO: 1) with C-14 sterol reductase of *Saccharomyces cerevisiae* (Erg24) and *Schizosaccharomyces pombe* (Pombe), and C-24 sterol reductase of *Sz. pombe* (Sts1) and *S. cerevisiae* (Ygl022).

Fig. 12 is a schematic illustration showing that the predicted *ELL* amino acid sequence (designated Ath; SEQ ID NO: 1) shares homology to human and chicken lamin B receptor.

Figs. 13A-13B are photographs showing that the *ell* phenotype was not corrected by exogenous feeding of brassinolide (1 μ M) in either dark (Fig. 13A) or light (Fig. 13B). From left to right in Figs. 13A-13B: wild-type; *ell*; wild-type + brassinolide; and *ell* + brassinolide.

- 12 -

Fig. 14 is a schematic illustration showing the nucleotide sequence of an *Arabidopsis* C-14 sterol reductase (SEQ ID NO: 2) and its deduced amino acid sequence polypeptide (SEQ ID NO: 1).

Fig. 15 is a schematic illustration showing the genomic nucleotide sequence of an *Arabidopsis* C-14 sterol reductase polypeptide (SEQ ID NO: 3).

Fig. 16 is a schematic illustration showing the sequence comparison between the genomic nucleotide sequence (SEQ ID NO: 3) and cDNA sequences (SEQ ID NO: 2) of an *Arabidopsis* C-14 sterol reductase.

There now follows a description of an *Arabidopsis* mutant, *ell* (extra long life), that displays a life span that is at least three times greater than wild-type plants. The *ell* mutant was isolated by T-DNA tagging methods and was shown to encode a novel C-14 sterol reductase. This example is provided for the purpose of illustrating the invention, and should not be construed as limiting.

Identification and Developmental Effects of the *ell* Mutation By screening for mutants displaying BR deficiency or constitutive cytokinin activity, a recessive mutation causing pleiotropic developmental effects was identified according to conventional methods in an *Arabidopsis* T-DNA insertional mutant collection (Feldmann, *Plant J.* 1:71, 1991; Errampalli et al., *Plant Cell* 3: 149, 1991). This mutant, termed "ell", was found to have a number of developmental abnormalities.

For example, unlike wild-type plants, *ell* mutants displayed constitutive light-morphogenesis (Fig. 2), similar to the *Arabidopsis det2* (Chory et al., *Plant Cell* 3: 445, 1991) and *cpd* (Szekeres et al., *Cell* 85: 171, 1996) mutants. In addition, compared to wild-type plants, *ell* plants had darker green rosette leaves (Figs. 3A-3B), reduced apical dominance (Fig. 4), stunted hairy roots, and irregular hypocotyl and cotyledons (Fig. 5). Furthermore, as shown in Figs. 6A-6B, the *ell* mutant showed reduced and ruffled sepals and petals. The *ell* mutant also showed delayed and altered

embryo development (Figs. 7A-7F) and was found to have reduced fertility, producing wrinkled seeds that precociously germinated (Figs. 8A-8B). In addition, the various phenotypes of *ell* overlapped with *amp-1* (*pt1*) (Chaudhury et. al., *Plant J.* 4: 907, 1993) and *hauptling* (Jurgens et al., *Ann. Rev. Genet.* 28: 351, 1994), including
5 supernumerary cotyledons (Figs. 9A-9F).

Finally, as shown in Figs. 1A-1B, the morphology of the T-DNA tagged *ell* mutant seedlings was phenocopied by treating wild-type seedlings with 30 μ M dimethylallylamine purine (2ip), a synthetic cytokinin.

Despite having a number of developmental abnormalities, *ell* mutants were
10 found to have a life span that was at least three times greater than wild-type plants.

Genetic Analysis and Molecular Cloning of ELL

Standard segregation analysis indicated that *ell* is a recessive mutation. The T2 population of the transgenic line carrying the *ell* mutant showed a 3:1 Mendelian segregation of the T-DNA using kanamycin resistance (*kan*^r) as a selectable marker.
15 Of the *kan*^r plants, thirty-three percent showed the *ell* phenotype, indicating that the *ell* mutation was recessive. A T3 population was then generated from selfed T2 *kan*^r plants having the wild-type phenotype, and the *kan*^r marker showed a 3:1 segregation. Of the seventy-five percent displaying *kan*^r, twenty-five percent showed the *ell* phenotype. Because *ell* homozygous plants were found to be either lethal or sterile,
20 T2 heterozygous *ell* plants were subsequently backcrossed to wild-type plants for additional segregation analysis. The resulting F1 population from this backcross showed a 1:1 segregation of the *kan*^r marker; no plants were observed having the *ell* phenotype. The F1 *kan*^r individuals of the backcross were then selfed to produce an F2 population. Seventy-five percent of this F2 population was found to be *kan*^r, and
25 thirty-three percent of the *kan*^r resistant plants showed the *ell* phenotype, confirming the recessive nature of this mutant. Consistent segregation of the *ell* phenotype and *kan*^r marker was also observed in a subsequent backcross, further indicating that *ell* was tagged by the T-DNA.

- 14 -

Genomic DNA blot analysis, using an NPTII probe derived from the T-DNA vector, showed that a unique single copy of T-DNA was integrated into the *ell* genome. This result, together with the segregation data described above, further indicated that the *ell* phenotype was associated with the *kan'* marker, and that the *ell* mutation resulted from a single T-DNA insertion in the *Arabidopsis* genome.

The T-DNA-tagged locus was then isolated by constructing a genomic DNA library from the *ell* mutant and was mapped by hybridization using the NPTII probe. Fig. 10A shows the physical map of the T-DNA tagged locus that was determined by DNA hybridization. One of three genomic clones that were found to hybridize to the NPTII probe was partially sequenced and found to have a complete T-DNA insertion and flanking plant sequences. A segment of this genomic clone containing both T-DNA and plant sequences was then used to screen a genomic library that was prepared from wild-type plants. Two positive clones that were identified in this screen were then sequenced. The genomic nucleotide sequence is presented in Fig. 15 (SEQ ID NO: 3).

The T-DNA-plant DNA insert junctions were also used as probes to screen a cDNA library that was prepared from wild-type plants. One isolated cDNA clone, designated D13, was found to have a nucleotide sequence (SEQ ID NO: 1) that matched the genomic sequences flanking the right T-DNA border. Comparison of the cDNA (Fig. 14) with the genomic DNA sequence (Fig. 15) also revealed that the T-DNA was inserted at a location forty base pairs upstream of the 5' end of the *ELL* cDNA transcript (Fig. 16). The complete genomic fragment covering the cDNA sequence was composed of 14 exons and 13 introns (Fig. 10A). Probes that were prepared from both the cDNA or genomic clone were then used for DNA blot analysis. Results from this analysis confirmed that the *ELL* gene was of plant origin.

We also determined the chromosomal position of *ELL* by standard segregation analysis of restriction fragment length polymorphisms (RFLPs) in recombinant inbred lines (Nam et al., *Plant Cell* 1: 699, 1989; Lister and Dean, *Plant J.* 4: 745, 1993;

Hauge et al., *Plant J.* 3: 745, 1993; Schmidt et al., *Science* 270: 480, 1995; Zachgo et al., *Genomic Res.* 6: 19, 1996). By this analysis, we found that *ELL* is located on chromosome 3 and is flanked by the chromosomal markers by mi456 and g2778 (Fig. 10B).

5

ELL Encodes a Novel C-14 Sterol Reductase

A comparison of the deduced polypeptide sequence of the full-length *ELL* cDNA clone to the GenBank database showed that *ELL* had 35% identity to C-14 sterol reductase (Erg24) in yeast (Lorenz and Parks, *DNA Cell Biol.* 9: 685, 1992; Lai et al., *Gene* 140: 41, 1994) (Fig. 11) and 40% identity to the lamin B receptor (LBR) in humans (Ye and Worman, *J. Biol. Chem.* 269: 11306, 1994) (Fig. 12). In addition, the amino acid sequence of *ELL* predicted several hydrophobic regions and between eight to nine transmembrane domains, consistent with the yeast Erg24 and human LBR. However, *ELL* was observed to lack a basic nucleoplasmic amino-terminal domain of about 200 amino acids that has been identified in human LBR. Database searches also revealed that at least two *Arabidopsis* expression sequence tagged (EST) clones (GenBank accession numbers T45011 and T42407) shared homology to *ELL*. DNA sequencing revealed that T45011 encodes an unknown gene with 60% nucleotide sequence identity to *ELL*. The predicted amino acid sequence of T45011 was also observed to have greater than 50% identity to the yeast ERG24 and human LBR. These results further confirmed that *ELL* is encoded by a gene that is a member of the C-14 sterol reductase gene family. T42407 was found to encode an *Arabidopsis* sterol $\Delta 7$ -reductase (Lecain et al., *J. Biol. Chem.* 271: 10866, 1996) that shares 32% amino acid identity to *ELL*.

RNA blot analysis indicated multiple transcripts hybridizing to the full-length *ELL* cDNA.

To determine whether the *ell* mutant phenotype is corrected by exogenous feeding of brassinolide, we germinated *ell* seedlings on agar plates containing 1 μ M

- 16 -

brassinolide or 1 μ M 24-epibrassinolide (Li et al., *Science* 272: 398, 1996). The results of these experiments showed that the presence of brassinolide or 24-epibrassinolide, in the growth medium of *ell* plants did not alter the mutant phenotype (Fig. 13A-13B). Thus, it appears that steroid compounds other than BRs are needed to restore an *ell* mutant to a normal growth and development phenotype, as reflected by the pleiotropic phenotypes such as stunted roots (Fig. 2) and impaired embryogenesis (Fig. 7A-7F).

To confirm that ELL activity was upstream of DET2 in the sterol biosynthesis pathway, a double mutant between *ell* and *det2* was constructed and analyzed. The phenotype of *det2/ell* was indistinguishable from *ell*, further supporting the hypothesis that DET2 was epistatic to ELL.

Isolation of Other C-14 Sterol Reductase cDNAs and Genomic DNAs

Based on the C-14 sterol reductase genes and polypeptides described herein, the isolation of additional plant C-14 sterol reductase coding sequences is made possible using standard strategies and techniques that are well known in the art. For example, using all or a portion of the amino acid sequence of a C-14 sterol reductase polypeptide, one may readily design C-14 sterol reductase-specific oligonucleotide probes, including C-14 sterol reductase degenerate oligonucleotide probes (i.e., a mixture of all possible coding sequences for a given amino acid sequence). These oligonucleotides may be based upon the sequence of either DNA strand and any appropriate portion of the C-14 sterol reductase sequence (for example, Fig. 14; SEQ ID NOS: 2 and 1, respectively; and Fig. 15 (SEQ ID NO: 3). General methods for designing and preparing such probes are provided, for example, in Ausubel et al., 1996, *Current Protocols in Molecular Biology*, Wiley Interscience, New York, and Berger and Kimmel, *Guide to Molecular Cloning Techniques*, 1987, Academic Press, New York. These oligonucleotides are useful for C-14 sterol reductase gene isolation, either through their use as probes capable of hybridizing to C-14 sterol reductase

- 17 -

complementary sequences or as primers for various amplification techniques, for example, polymerase chain reaction (PCR) cloning strategies.

Hybridization techniques and screening procedures are well known to those skilled in the art and are described, for example, in Ausubel et al. (*supra*); Berger and Kimmel (*supra*); and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York. If desired, a combination of different oligonucleotide probes may be used for the screening of a recombinant DNA library. The oligonucleotides may be detectably-labeled using methods known in the art and used to probe filter replicas from a recombinant DNA library. Recombinant DNA libraries are prepared according to methods well known in the art, for example, as described in Ausubel et al. (*supra*), or they may be obtained from commercial sources.

For detection or isolation of closely related C-14 sterol reductase sequences having greater than 80% identity, high stringency conditions are preferably used; such conditions include hybridization at about 65°C and about 50% formamide, a first wash at about 65°C, about 2X SSC, and 1% SDS, followed by a second wash at about 65°C and about 0.1% SDS, and 0.1X SSC. Lower stringency conditions for detecting C-14 sterol reductase genes having about 40-50% sequence identity to the C-14 sterol reductase genes described herein include, for example, hybridization at about 37°C in the absence of formamide, a first wash at about 37°C, about 6X SSC, and about 1% SDS, and a second wash at about 37°C, about 6X SSC, and about 1% SDS. These stringency conditions are exemplary; other appropriate conditions may be determined by those skilled in the art.

As discussed above, C-14 sterol reductase oligonucleotides may also be used as primers in amplification cloning strategies, for example, using PCR. PCR methods are well known in the art and are described, for example, in *PCR Technology*, Erlich, ed., Stockton Press, London, 1989; *PCR Protocols: A Guide to Methods and Applications*, Innis et al., eds., Academic Press, Inc., New York, 1990; and Ausubel et

- 18 -

al. (*supra*). Primers are optionally designed to allow cloning of the amplified product into a suitable vector, for example, by including appropriate restriction sites at the 5' and 3' ends of the amplified fragment (as described herein). If desired, C-14 sterol reductase sequences may be isolated using the PCR "RACE" technique, or Rapid Amplification of cDNA Ends (see, e.g., Innis et al. (*supra*)). By this method, oligonucleotide primers based on a C-14 sterol reductase sequence are oriented in the 3' and 5' directions and are used to generate overlapping PCR fragments. These overlapping 3'- and 5'-end RACE products are combined to produce an intact full-length cDNA. This method is described in Innis et al. (*supra*); and Frohman et al., *Proc. Natl. Acad. Sci. USA* 85: 8998, 1988.

Alternatively, any plant cDNA expression library may be screened by functional complementation of a yeast C-14 reductase mutant (for example, the *erg24* mutant described by Lorenz and Parks, *DNA Cell Biol.* 9: 685, 1992) according to standard methods.

Useful C-14 sterol reductase sequences may be isolated from any appropriate organism. Confirmation of a sequence's relatedness to the C-14 sterol reductase polypeptide family may be accomplished by a variety of conventional methods including, but not limited to, functional complementation assays and sequence comparison. In addition, the activity of any C-14 sterol reductase sequence may be evaluated according to any of the techniques described herein.

C-14 Sterol Reductase Polypeptide Expression

C-14 sterol reductase polypeptides may be produced by transformation of a suitable host cell with all or part of a C-14 sterol reductase cDNA (for example, the cDNA described above) in a suitable expression vehicle or with a plasmid construct engineered for increasing the expression of a C-14 sterol reductase polypeptide (*supra*) *in vivo*.

- 19 -

Those skilled in the field of molecular biology will understand that any of a wide variety of expression systems may be used to provide the recombinant protein. The precise host cell used is not critical to the invention. The C-14 sterol reductase protein may be produced in a prokaryotic host, for example, *E. coli*, or in a eukaryotic host, for example, *Saccharomyces cerevisiae*, mammalian cells (for example, COS 1 or NIH 3T3 cells), or any of a number of plant cells including, without limitation, algae, tree species, ornamental species, temperate fruit species, tropical fruit species, vegetable species, legume species, monocots, dicots, or in any plant of commercial or agricultural significance. Particular examples of suitable plant hosts include, but are not limited to, Conifers, Petunia, Tomato, Potato, Tobacco, *Arabidopsis*, Lettuce, Sunflower, Oilseed rape, Flax, Cotton, Sugarbeet, Celery, Soybean, Alfalfa, *Medicago*, Lotus, *Vigna*, Cucumber, Carrot, Eggplant, Cauliflower, Horseradish, Morning Glory, Poplar, Walnut, Apple, Asparagus, Rice, Maize, Millet, Onion, Barley, Orchard grass, Oat, Rye, and Wheat.

Such cells are available from a wide range of sources including the American Type Culture Collection (Rockland, MD); or from any of a number seed companies, for example, W. Atlee Burpee Seed Co. (Warminster, PA), Park Seed Co. (Greenwood, SC), Johnny Seed Co. (Albion, ME), or Northrup King Seeds (Harstville, SC). Descriptions and sources of useful host cells are also found in Vasil I.K., *Cell Culture and Somatic Cell Genetics of Plants*, Vol I, II, III Laboratory Procedures and Their Applications Academic Press, New York, 1984; Dixon, R.A., *Plant Cell Culture-A Practical Approach*, IRL Press, Oxford University, 1985; Green et al., *Plant Tissue and Cell Culture*, Academic Press, New York, 1987; and Gasser and Fraley, *Science* 244: 1293, 1989.

For prokaryotic expression, DNA encoding a C-14 sterol reductase polypeptide is carried on a vector operably linked to control signals capable of effecting expression in the prokaryotic host. If desired, the coding sequence may contain, at its 5' end, a sequence encoding any of the known signal sequences capable of effecting secretion

- 20 -

of the expressed protein into the periplasmic space of the host cell, thereby facilitating recovery of the protein and subsequent purification. Prokaryotes most frequently used are various strains of *E. coli*; however, other microbial strains may also be used.

Plasmid vectors are used which contain replication origins, selectable markers, and control sequences derived from a species compatible with the microbial host.

Examples of such vectors are found in Pouwels et al. (*supra*) or Ausubel et al. (*supra*).

Commonly used prokaryotic control sequences (also referred to as "regulatory elements") are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences. Promoters commonly used to direct protein expression include the beta-lactamase (penicillinase), the lactose (*lac*) (Chang et al., *Nature* 198: 1056, 1977), the tryptophan (*Trp*) (Goeddel et al., *Nucl. Acids Res.* 8: 4057, 1980), and the *tac* promoter systems, as well as the lambda-derived *P_L* promoter and N-gene ribosome binding site (Simatake et al., *Nature* 292: 128, 1981).

One particular bacterial expression system for C-14 sterol reductase polypeptide production is the *E. coli* pET expression system (Novagen, Inc., Madison, WI). According to this expression system, DNA encoding a C-14 sterol reductase polypeptide is inserted into a pET vector in an orientation designed to allow expression. Since the C-14 sterol reductase gene is under the control of the T7 regulatory signals, expression of C-14 sterol reductase is induced by inducing the expression of T7 RNA polymerase in the host cell. This is typically achieved using host strains which express T7 RNA polymerase in response to IPTG induction. Once produced, recombinant C-14 sterol reductase polypeptide is then isolated according to standard methods known in the art, for example, those described herein.

Another bacterial expression system for C-14 sterol reductase polypeptide production is the pGEX expression system (Pharmacia). This system employs a GST gene fusion system which is designed for high-level expression of genes or gene fragments as fusion proteins with rapid purification and recovery of functional gene

products. The protein of interest is fused to the carboxyl terminus of the glutathione S-transferase protein from *Schistosoma japonicum* and is readily purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4B. Fusion proteins can be recovered under mild conditions by elution with glutathione.

5 Cleavage of the glutathione S-transferase domain from the fusion protein is facilitated by the presence of recognition sites for site-specific proteases upstream of this domain. For example, proteins expressed in pGEX-2T plasmids may be cleaved with thrombin; those expressed in pGEX-3X may be cleaved with factor Xa.

For eukaryotic expression, the method of transformation or transfection and the
10 choice of vehicle for expression of the C-14 sterol reductase polypeptide will depend on the host system selected. Transformation and transfection methods are described, e.g., in Ausubel et al. (*supra*); Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989; Gelvin et al., *Plant Molecular Biology Manual*, Kluwer Academic Publishers, 1990; Kindle, K., *Proc. Natl. Acad. Sci., U.S.A*
15 87: 1228, 1990; Potrykus, I., *Annu. Rev. Plant Physiol. Plant Mol. Biology* 42: 205, 1991; and BioRad (Hercules, CA) Technical Bulletin #1687 (Biolistic Particle Delivery Systems). Expression vehicles may be chosen from those provided, e.g., in *Cloning Vectors: A Laboratory Manual* (P.H. Pouwels et al., 1985, Supp. 1987); Gasser and Fraley (*supra*); Clontech Molecular Biology Catalog (Catalog 1992/93
20 Tools for the Molecular Biologist, Palo Alto, CA); and the references cited above.

Most preferably, an C-14 sterol reductase polypeptide is produced by a stably-transfected plant cell line, a transiently-transfected plant cell line, or by a transgenic plant. A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants are available to the public; such vectors are
25 described in Pouwels et al. (*supra*), Weissbach and Weissbach (*supra*), and Gelvin et al. (*supra*). Methods for constructing such cell lines are described in, e.g., Weissbach and Weissbach (*supra*), and Gelvin et al. (*supra*). Typically, plant expression vectors include (1) a cloned plant gene under the transcriptional control of 5' and 3' regulatory

sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (for example, one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Alternatively, the C-14 sterol reductase polypeptide may be produced using a transient expression system (e.g., the maize transient expression system described by Sheen, *Plant Cell* 2: 1027, 1990).

Once the desired C-14 sterol reductase nucleic acid sequences is obtained, it may be manipulated in a variety of ways known in the art. For example, where the sequence involves non-coding flanking regions, the flanking regions may be subjected to mutagenesis.

The C-14 sterol reductase DNA sequence of the invention may, if desired, be combined with other DNA sequences in a variety of ways. The C-14 sterol reductase DNA sequence of the invention may be employed with all or part of the gene sequences normally associated with the C-14 sterol reductase protein. In its component parts, a DNA sequence encoding a C-14 sterol reductase protein is combined in a DNA construct having a transcription initiation control region capable of promoting transcription and translation in a host cell.

In general, the constructs will involve regulatory regions functional in plants which provide for modified production of C-14 sterol reductase protein as discussed herein. The open reading frame coding for the C-14 sterol reductase protein or functional fragment thereof will be joined at its 5' end to a transcription initiation regulatory region such as the sequence naturally found in the 5' upstream region of the C-14 sterol reductase structural gene. Numerous other transcription initiation regions are available which provide for constitutive or inducible regulation.

- 23 -

For applications where developmental, cell, tissue, hormonal, or environmental expression is desired, appropriate 5' upstream non-coding regions are obtained from other genes, for example, from genes regulated during meristem development, seed development, embryo development, or leaf development.

5 Regulatory transcript termination regions may also be provided in DNA constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the C-14 sterol reductase protein or any convenient transcription termination region derived from a different gene source. The transcript termination region will contain preferably at least 1-3 kb of sequence 3' to the
10 structural gene from which the termination region is derived. Plant expression constructs having C-14 sterol reductase as the DNA sequence of interest for expression (in either the sense or antisense orientation) may be employed with a wide variety of plant life, particularly plant life involved in the production of storage reserves (for example, those involving carbon and nitrogen metabolism). Such
15 genetically-engineered plants are useful for a variety of industrial and agricultural applications as discussed below. Importantly, this invention is applicable to dicotyledons and monocotyledons, and will be readily applicable to any new or improved transformation or regeneration method.

 An example of a useful plant promoter according to the invention is a
20 caulimovirus promoter, for example, a cauliflower mosaic virus (CaMV) promoter. These promoters confer high levels of expression in most plant tissues, and the activity of these promoters is not dependent on virally encoded proteins. CaMV is a source for both the 35S and 19S promoters. In most tissues of transgenic plants, the CaMV 35S promoter is a strong promoter (see, e.g., Odell et al., *Nature* 313: 810 1985). The
25 CaMV promoter is also highly active in monocots (see, e.g., Dekeyser et al., *Plant Cell* 2: 591, 1990; Terada and Shimamoto, *Mol. Gen. Genet.* 220: 389, 1990). Moreover, activity of this promoter can be further increased (i.e., between 2-10 fold) by duplication of the CaMV 35S promoter (see e.g., Kay et al., *Science* 236: 1299,

1987; Ow et al., *Proc. Natl. Acad. Sci., U.S.A.* 84: 4870, 1987; and Fang et al., *Plant Cell* 1: 141, 1989).

Other useful plant promoters include, without limitation, the nopaline synthase promoter (An et al., *Plant Physiol.* 88: 547, 1988) and the octopine synthase promoter (Fromm et al., *Plant Cell* 1: 977, 1989).

For certain applications, it may be desirable to produce the C-14 sterol reductase gene product in an appropriate tissue, at an appropriate level, or at an appropriate developmental time. For this purpose, there are an assortment of gene promoters, each with its own distinct characteristics embodied in its regulatory sequences, shown to be regulated in response to the environment, hormones, and/or developmental cues. These include gene promoters that are responsible for heat-regulated gene expression (see, e.g., Callis et al., *Plant Physiol.* 88: 965, 1988; Takahashi and Komeda, *Mol. Gen. Genet.* 219: 365, 1989; and Takahashi et al. *Plant J.* 2: 751, 1992), light-regulated gene expression (e.g., the pea *rbcS-3A* described by Kuhlemeier et al., *Plant Cell* 1: 471, 1989; the maize *rbcS* promoter described by Schäffner and Sheen, *Plant Cell* 3: 997, 1991; or the chlorophyll a/b-binding protein gene found in pea described by Simpson et al., *EMBO J.* 4: 2723, 1985), hormone-regulated gene expression (for example, the abscisic acid (ABA) responsive sequences from the *Em* gene of wheat described by Marcotte et al., *Plant Cell* 1: 969, 1989; the ABA-inducible HVA1 and HVA22, and rd29A promoters described for barley and *Arabidopsis* by Straub et al., *Plant Cell* 6: 617, 1994, Shen et al., *Plant Cell* 7: 295, 1995; and wound-induced gene expression (for example, of *wun1* described by Siebertz et al., *Plant Cell* 1: 961, 1989), or organ-specific gene expression (for example, of the tuber-specific storage protein gene described by Roshal et al., *EMBO J.* 6: 1155, 1987; the 23-kDa zein gene from maize described by Schernthaner et al., *EMBO J.* 7: 1249, 1988; or the French bean β -phaseolin gene described by Bustos et al., *Plant Cell* 1: 839, 1989).

Plant expression vectors may also optionally include RNA processing signals, e.g., introns, which have been shown to be important for efficient RNA synthesis and accumulation (Callis et al., *Genes and Dev.* 1: 1183, 1987). The location of the RNA splice sequences can dramatically influence the level of transgene expression in plants.

5 In view of this fact, an intron may be positioned upstream or downstream of a C-14 sterol reductase polypeptide-encoding sequence in the transgene to modulate levels of gene expression.

In addition to the aforementioned 5' regulatory control sequences, the expression vectors may also include regulatory control regions which are generally present in the 3' regions of plant genes (Thornburg et al., *Proc. Natl. Acad. Sci. U.S.A.* 84: 744, 1987; An et al., *Plant Cell* 1: 115, 1989). For example, the 3' terminator region may be included in the expression vector to increase stability of the mRNA. One such terminator region may be derived from the PI-II terminator region of potato.

10 In addition, other commonly used terminators are derived from the octopine or nopaline synthase signals.

The plant expression vector also typically contains a dominant selectable marker gene used to identify those cells that have become transformed. Useful selectable genes for plant systems include genes encoding antibiotic resistance genes, for example, those encoding resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin, or spectinomycin. Genes required for photosynthesis may also be used as selectable markers in photosynthetic-deficient strains. Alternatively, the green-fluorescent protein from the jellyfish *Aequorea victoria* may be used as a selectable marker (Sheen et al., *Plant J.* 8:777, 1995; Chiu et al., *Current Biology* 6: 325, 1996). Finally, genes encoding herbicide resistance may be used as selectable markers; useful herbicide resistance genes include the *bar* gene encoding the enzyme phosphinothricin acetyltransferase and conferring resistance to the broad spectrum herbicide Basta® (Hoechst AG, Frankfurt, Germany).

20

25

Efficient use of selectable markers is facilitated by a determination of the susceptibility of a plant cell to a particular selectable agent and a determination of the concentration of this agent which effectively kills most, if not all, of the transformed cells. Some useful concentrations of antibiotics for tobacco transformation include, e.g., 75-100 µg/ml (kanamycin), 20-50 µg/ml (hygromycin), or 5-10 µg/ml (bleomycin). A useful strategy for selection of transformants for herbicide resistance is described, e.g., by Vasil et al., *supra*.

It should be readily apparent to one skilled in the art of molecular biology, especially in the field of plant molecular biology, that the level of gene expression is dependent, not only on the combination of promoters, RNA processing signals, and terminator elements, but also on how these elements are used to increase the levels of selectable marker gene expression.

Plant Transformation

Upon construction of the plant expression vector, several standard methods are available for introduction of the vector into a plant host, thereby generating a transgenic plant. These methods include (1) *Agrobacterium*-mediated transformation (*A. tumefaciens* or *A. rhizogenes*) (see, e.g., Lichtenstein and Fuller In: *Genetic Engineering*, vol 6, PWJ Rigby, ed, London, Academic Press, 1987; and Lichtenstein, C.P., and Draper, J., In: *DNA Cloning*, Vol II, D.M. Glover, ed, Oxford, IRI Press, 1985)), (2) the particle delivery system (see, e.g., Gordon-Kamm et al., *Plant Cell* 2: 603 (1990); or BioRad Technical Bulletin 1687, *supra*), (3) microinjection protocols (see, e.g., Green et al., *supra*), (4) polyethylene glycol (PEG) procedures (see, e.g., Draper et al., *Plant Cell Physiol.* 23: 451, 1982; or e.g., Zhang and Wu, *Theor. Appl. Genet.* 76: 835, 1988), (5) liposome-mediated DNA uptake (see, e.g., Freeman et al., *Plant Cell Physiol.* 25: 1353, 1984), (6) electroporation protocols (see, e.g., Gelvin et al., *supra*; Dekeyser et al., *supra*; Fromm et al., *Nature* 319: 791, 1986; Sheen *Plant Cell* 2: 1027, 1990; or Jang and Sheen *Plant Cell* 6: 1665, 1994), and (7) the vortexing

method (see, e.g., Kindle *supra*). The method of transformation is not critical to the invention. Any method which provides for efficient transformation may be employed. As newer methods are available to transform crops or other host cells, they may be directly applied.

5 The following is an example outlining one particular technique, an *Agrobacterium*-mediated plant transformation. By this technique, the general process for manipulating genes to be transferred into the genome of plant cells is carried out in two phases. First, cloning and DNA modification steps are carried out in *E. coli*, and the plasmid containing the gene construct of interest is transferred by conjugation or
10 electroporation into *Agrobacterium*. Second, the resulting *Agrobacterium* strain is used to transform plant cells. Thus, for the generalized plant expression vector, the plasmid contains an origin of replication that allows it to replicate in *Agrobacterium* and a high copy number origin of replication functional in *E. coli*. This permits facile production and testing of transgenes in *E. coli* prior to transfer to *Agrobacterium* for
15 subsequent introduction into plants. Resistance genes can be carried on the vector, one for selection in bacteria, for example, streptomycin, and another that will function in plants, for example, a gene encoding kanamycin resistance or herbicide resistance. Also present on the vector are restriction endonuclease sites for the addition of one or more transgenes and directional T-DNA border sequences which, when recognized by
20 the transfer functions of *Agrobacterium*, delimit the DNA region that will be transferred to the plant.

 In another example, plant cells may be transformed by shooting into the cell tungsten microprojectiles on which cloned DNA is precipitated. In the Biolistic Apparatus (Bio-Rad) used for the shooting, a gunpowder charge (22 caliber Power
25 Piston Tool Charge) or an air-driven blast drives a plastic macroprojectile through a gun barrel. An aliquot of a suspension of tungsten particles on which DNA has been precipitated is placed on the front of the plastic macroprojectile. The latter is fired at an acrylic stopping plate that has a hole through it that is too small for the

- 28 -

macroprojectile to pass through. As a result, the plastic macroprojectile smashes against the stopping plate, and the tungsten microprojectiles continue toward their target through the hole in the plate. For the instant invention the target can be any plant cell, tissue, seed, or embryo. The DNA introduced into the cell on the microprojectiles becomes integrated into either the nucleus or the chloroplast.

In general, transfer and expression of transgenes in plant cells are now routine practices to those skilled in the art, and have become major tools to carry out gene expression studies in plants and to produce improved plant varieties of agricultural or commercial interest.

Transgenic Plant Regeneration

Plant cells transformed with a plant expression vector can be regenerated, for example, from single cells, callus tissue, or leaf discs according to standard plant tissue culture techniques. It is well known in the art that various cells, tissues, and organs from almost any plant can be successfully cultured to regenerate an entire plant; such techniques are described, e.g., in Vasil *supra*; Green et al., *supra*; Weissbach and Weissbach, *supra*; and Gelvin et al., *supra*.

In one particular example, a cloned C-14 sterol reductase polypeptide or an antisense construct under the control of the 35S CaMV promoter and the nopaline synthase terminator and carrying a selectable marker (for example, kanamycin resistance) is transformed into *Agrobacterium*. Transformation of leaf discs (for example, of tobacco leaf discs), with vector-containing *Agrobacterium* is carried out as described by Horsch et al. (*Science* 227: 1229, 1985). Putative transformants are selected after a few weeks (for example, 3 to 5 weeks) on plant tissue culture media containing kanamycin (e.g. 100 µg/ml). Kanamycin-resistant shoots are then placed on plant tissue culture media without hormones for root initiation. Kanamycin-resistant plants are then selected for greenhouse growth. If desired, seeds from self-fertilized transgenic plants can then be sowed in a soil-less medium and grown in a

- 29 -

greenhouse. Kanamycin-resistant progeny are selected by sowing surfaced sterilized seeds on hormone-free kanamycin-containing media. Analysis for the integration of the transgene is accomplished by standard techniques (see, for example, Ausubel et al. *supra*; Gelvin et al. *supra*).

5 Transgenic plants expressing the selectable marker are then screened for transmission of the transgene DNA by standard immunoblot and DNA detection techniques. Each positive transgenic plant and its transgenic progeny are unique in comparison to other transgenic plants established with the same transgene. Integration of the transgene DNA into the plant genomic DNA is in most cases random, and the
10 site of integration can profoundly affect the levels and the tissue and developmental patterns of transgene expression. Consequently, a number of transgenic lines are usually screened for each transgene to identify and select plants with the most appropriate expression profiles.

Transgenic lines are evaluated for levels of transgene expression. Expression at
15 the RNA level is determined initially to identify and quantitate expression-positive plants. Standard techniques for RNA analysis are employed and include PCR amplification assays using oligonucleotide primers designed to amplify only transgene RNA templates and solution hybridization assays using transgene-specific probes (see, e.g., Ausubel et al., *supra*). The RNA-positive plants are then analyzed for protein
20 expression by Western immunoblot analysis using C-14 sterol reductase specific antibodies (see, e.g., Ausubel et al., *supra*). In addition, *in situ* hybridization and immunocytochemistry according to standard protocols can be done using transgene-specific nucleotide probes and antibodies, respectively, to localize sites of expression within transgenic tissue.

25 Once the recombinant C-14 sterol reductase protein is expressed in any cell or in a transgenic plant (for example, as described above), it may be isolated, e.g., using affinity chromatography. In one example, an anti-C14 sterol reductase antibody (e.g., produced as described in Ausubel et al., *supra*, or by any standard technique) may be

- 30 -

attached to a column and used to isolate the polypeptide. Lysis and fractionation of C-14 sterol reductase-producing cells prior to affinity chromatography may be performed by standard methods (see, e.g., Ausubel et al., *supra*). Once isolated, the recombinant protein can, if desired, be further purified, for example, by high performance liquid chromatography (see, e.g., Fisher, *Laboratory Techniques In Biochemistry And Molecular Biology*, eds., Work and Burdon, Elsevier, 1980).

Polypeptides of the invention, particularly short C-14 sterol reductase protein fragments, can also be produced by chemical synthesis (e.g., by the methods described in *Solid Phase Peptide Synthesis*, 2nd ed., 1984 The Pierce Chemical Co., Rockford, IL).

These general techniques of polypeptide expression and purification can also be used to produce and isolate useful C-14 sterol reductase fragments or analogs.

Antibodies

C-14 sterol reductases described herein (or immunogenic fragments or analogs) may be used to raise antibodies useful in the invention; such polypeptides may be produced by recombinant or peptide synthetic techniques (see, e.g., *Solid Phase Peptide Synthesis*, 2nd ed., 1984, Pierce Chemical Co., Rockford, IL; Ausubel et al., *supra*). The peptides may be coupled to a carrier protein, such as KLH as described in Ausubel et al, *supra*. The KLH-peptide is mixed with Freund's adjuvant and injected into guinea pigs, rats, or preferably rabbits. Antibodies may be purified by peptide antigen affinity chromatography.

Monoclonal antibodies may be prepared using the C-14 sterol reductase polypeptides described above and standard hybridoma technology (see, e.g., Kohler et al., *Nature* 256:495, 1975; Kohler et al., *Eur. J. Immunol.* 6:511, 1976; Kohler et al., *Eur. J. Immunol.* 6:292, 1976; Hammerling et al., *In Monoclonal Antibodies and T Cell Hybridomas*, Elsevier, NY, 1981; Ausubel et al., *supra*).

Once produced, polyclonal or monoclonal antibodies are tested for specific C-14 sterol reductase recognition by Western blot or immunoprecipitation analysis (by the methods described in Ausubel et al., *supra*). Antibodies which specifically recognize C-14 sterol reductases are considered to be useful in the invention; such antibodies may be used, e.g., in an immunoassay to monitor the level of C-14 sterol reductase produced by a plant.

Use

Because the present invention provides for the genetic manipulation of a plant sterol biosynthetic pathway, this invention described is useful for a variety of agricultural and commercial purposes including, but not limited to, increasing crop yields, improving crop and ornamental quality, and reducing agricultural production costs. For example, the methods, DNA constructs, proteins, and transgenic plants described herein are useful for improving a number of fruit and vegetable characteristics including, but not limited to, texture, size, nutritional content, modification of sterol composition, disease and insect resistance, and ripening processes. In addition, genetic manipulation of plant sterol composition (for example, seed sterol composition) is useful for improving food quality and oil stability, and regulating the formation of compounds having anti-nutritional properties.

In one particular example, antisense C-14 sterol reductase sequences are useful for reducing the expression of C-14 sterol reductase expression in a transgenic plant. Such reduced expression of C-14 sterol reductase provides a means for increasing the life-span of such plants. Increased life-span extends reproductive period, delays senescence, and increases branch number for high productivity and yield. In addition, transgenic plants expressing antisense C-14 sterol reductase are useful for producing plants having reduced and more compact proportions. Such plants require less space and land requirements for their growth, and are more convenient and efficient to harvest.

- 32 -

Overproduction of the C-14 sterol reductase in transgenic plants is useful for enhancing the production of steroid compounds having a variety of medicinal or agricultural applications. For example, overproduction of mammalian steroid hormones in plants offers an inexpensive means for producing such hormones.

5 In addition, C-14 sterol reductase polypeptides disclosed herein are useful for the development of enzyme inhibitors of the sterol biosynthetic pathway.

Other Embodiments

10 In other embodiments, the invention includes any protein which is substantially identical to a crucifer C-14 sterol reductase polypeptide (Fig. 10; SEQ ID NO:1); such homologs include other substantially pure naturally-occurring plant C-14 sterol reductase proteins as well as allelic variants; natural mutants; induced mutants; proteins encoded by DNA that hybridizes to the C-14 sterol reductase DNA sequence of Fig. 14 (SEQ ID NO: 2) under high stringency conditions or, less preferably, under 15 low stringency conditions (e.g., washing at 2X SSC at 37°C with a probe length of at least 10-15 nucleotides), both as described herein; and proteins specifically bound by antisera directed to a C-14 sterol reductase polypeptide. The term also includes chimeric polypeptides that include a C-14 sterol reductase portion.

20 The invention further includes analogs of any naturally-occurring plant C-14 sterol reductase polypeptide. Analogs can differ from the naturally-occurring C-14 sterol reductase protein by amino acid sequence differences, by post-translational modifications, or by both. Analogs of the invention will generally exhibit at least 30%, more preferably 40%, and most preferably 50% or even 80-95% identity with all or part of a naturally-occurring plant C-14 sterol reductase amino acid sequence. The 25 length of sequence comparison is at least 15 amino acid residues, preferably at least 25 amino acid residues, and more preferably more than 35 amino acid residues. Modifications include *in vivo* and *in vitro* chemical derivatization of polypeptides, e.g., acetylation, carboxylation, phosphorylation, or glycosylation; such modifications

- 33 -

may occur during polypeptide synthesis or processing or following treatment with isolated modifying enzymes. Analogs can also differ from the naturally-occurring C-14 sterol reductase polypeptide by alterations in primary sequence. These include genetic variants, both natural and induced (for example, resulting from random mutagenesis by irradiation or exposure to ethanemethylsulfate or by site-specific mutagenesis as described in Sambrook, Fritsch and Maniatis, *Molecular Cloning: A Laboratory Manual* (2d ed.), CSH Press, 1989, or Ausubel et al., *supra*). Also included are cyclized peptides, molecules, and analogs which contain residues other than L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., β or γ amino acids.

In addition to full-length polypeptides, the invention also includes C-14 sterol reductase polypeptide fragments. As used herein, the term "fragment," means at least 20 contiguous amino acids, preferably at least 30 contiguous amino acids, more preferably at least 50 contiguous amino acids, and most preferably at least 60 to 80 or more contiguous amino acids. Fragments of C-14 sterol reductase polypeptides can be generated by methods known to those skilled in the art or may result from normal protein processing (e.g., removal of amino acids from the nascent polypeptide that are not required for biological activity or removal of amino acids by alternative mRNA splicing or alternative protein processing events).

Furthermore, the invention includes nucleotide sequences that facilitate specific detection of a C-14 sterol reductase nucleic acid. Thus, C-14 sterol reductase sequences described herein (e.g., SEQ ID NO: 2 and 3) or portions thereof may be used as probes to hybridize to nucleotide sequences from other plants (e.g., dicots, monocots, gymnosperms, and algae) by standard hybridization techniques under conventional conditions. Sequences that hybridize to a C-14 sterol reductase coding sequence or its complement and that encode a C-14 sterol reductase are considered useful in the invention. As used herein, the term "fragment," as applied to nucleic acid sequences, means at least 5 contiguous nucleotides, preferably at least 10 contiguous

- 34 -

nucleotides, more preferably at least 20 to 30 contiguous nucleotides, and most preferably at least 40 to 80 or more contiguous nucleotides. Fragments of C-14 sterol reductase nucleic acid sequences can be generated by methods known to those skilled in the art.

5 All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication or patent application was specifically and individually indicated to be incorporated by reference.

Other embodiments are within the following claims.

10

What is claimed is:

- 35 -

Claims

1. A substantially pure plant C-14 sterol reductase polypeptide.

5 2. A substantially pure polypeptide comprising an amino acid sequence substantially identical to the sequence shown in Fig. 14 (SEQ ID NO: 1).

 3. The polypeptide of claim 1 or 2, wherein said polypeptide comprises the amino acid sequence shown in Fig. 14 (SEQ ID NO: 1).

10 4. The polypeptide of claim 1 or 2, wherein said polypeptide is from a dicot.

 5. The polypeptide of claim 4, wherein said dicot is a crucifer.

15 6. A purified DNA encoding a C-14 sterol reductase polypeptide.

 7. A purified DNA comprising a sequence substantially identical to the DNA sequence shown in Fig. 14 (SEQ ID NO: 2).

20 8. The purified DNA of claim 6 or 7, wherein said DNA comprises the sequence shown in Fig. 14 (SEQ ID NO: 2).

 9. The purified DNA of claim 6 or 7, wherein said DNA encodes a polypeptide which has an amino acid sequence substantially identical to that shown in Fig. 14 (SEQ ID NO: 1).

25 10. The purified DNA of claim 9, wherein said DNA encodes a polypeptide which has the amino acid sequence shown in Fig. 14 (SEQ ID NO: 1).

- 36 -

11. The purified DNA of claim 6 or 7, wherein said DNA is from a dicot.

12. The purified DNA of claim 11, wherein said dicot is a crucifer.

5 13. A vector comprising the purified DNA of claim 6 or 7.

14. A cell comprising the purified DNA of claim 6 or 7.

10 15. A method of producing a recombinant C-14 sterol reductase polypeptide comprising

providing a cell transformed with purified DNA encoding a C-14 sterol reductase polypeptide positioned for expression in said cell, and culturing said transformed cell under conditions for expressing said DNA.

15 16. The method of claim 15, wherein said method further comprises recovering said recombinant C-14 sterol reductase polypeptide.

17. The method of claim 15, wherein said cell is a plant cell.

20 18. A recombinant C-14 sterol reductase produced by the method of claim 15.

19. An isolated antibody which specifically recognizes and binds a plant C-14 sterol reductase polypeptide.

25 20. A transgenic plant which contains DNA encoding a C-14 sterol reductase polypeptide integrated into the genome of said plant, wherein said DNA is expressed in said transgenic plant.

- 37 -

21. A transgenic plant which contains DNA encoding an amino acid sequence substantially identical to the sequence shown in Fig. 14 (SEQ ID NO: 1) integrated into the genome of said plant, wherein said DNA is expressed in said transgenic plant.

5

22. A seed from a transgenic plant of claim 20 or 21.

23. A cell from a transgenic plant of claim 20 or 21.

10

24. A method of detecting a C-14 sterol reductase gene in a plant cell comprising:

contacting the purified DNA of claim 6 or 7 or a portion thereof greater than about 12 nucleotides in length with a preparation of genomic DNA from said plant cell under hybridization conditions providing detection of DNA sequences having about 15 40% or greater sequence identity to SEQ ID NO: 2.

25. A method of isolating a C-14 sterol reductase gene or portion thereof, said method comprising

20

(a) amplifying said C-14 sterol reductase gene using oligonucleotide primers, wherein said primers each have regions of complementarity to opposite DNA strands in a region of SEQ ID NO: 2; and

(b) isolating said C-14 sterol reductase gene or portion thereof.

25

26. A method for reducing the level of a C-14 sterol reductase polypeptide in a transgenic plant cell, said method comprising expressing in a plant cell an antisense C-14 sterol reductase nucleic acid sequence.

- 38 -

27. The method of claim 26, wherein said antisense C-14 sterol reductase nucleic acid sequence is encoded by a transgene integrated into the genome of said transgenic plant cell.

5 28. The method of claim 26, wherein said C-14 sterol reductase sequence is SEQ ID NO: 2 or SEQ ID NO: 3.

 29. The method of claim 26, wherein said method further comprises growing a transgenic plant from said transgenic plant cell, whereby the level of the C-14 sterol
10 reductase polypeptide is reduced in said transgenic plant.

 30. A method for increasing the level of a C-14 sterol reductase in a transgenic plant cell, comprising expressing in said transgenic plant cell a nucleic acid sequence encoding a polypeptide substantially identical to SEQ ID NO: 1.
15

 31. A transgenic plant comprising purified DNA encoding a plant C-14 sterol reductase polypeptide, said DNA comprising a knockout mutation in said C-14 sterol reductase sequence.

FIGURE 1A

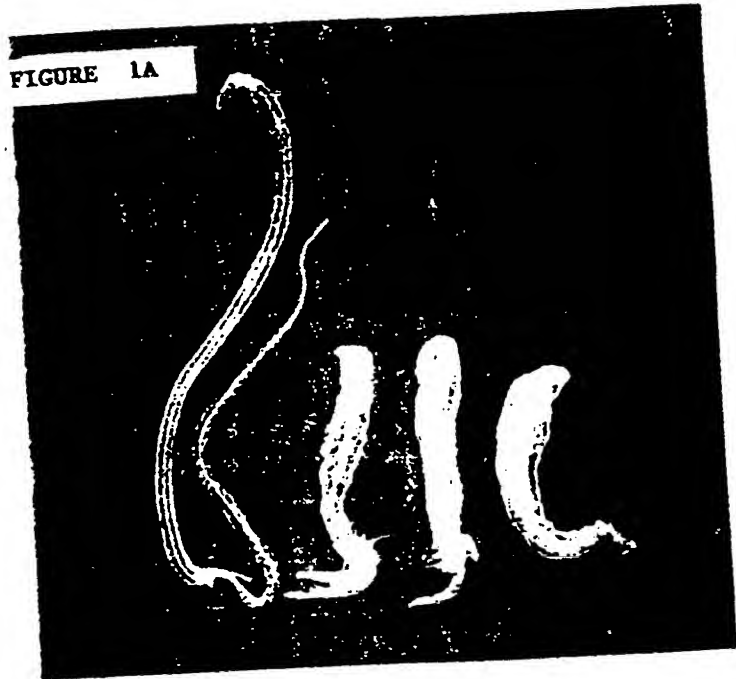
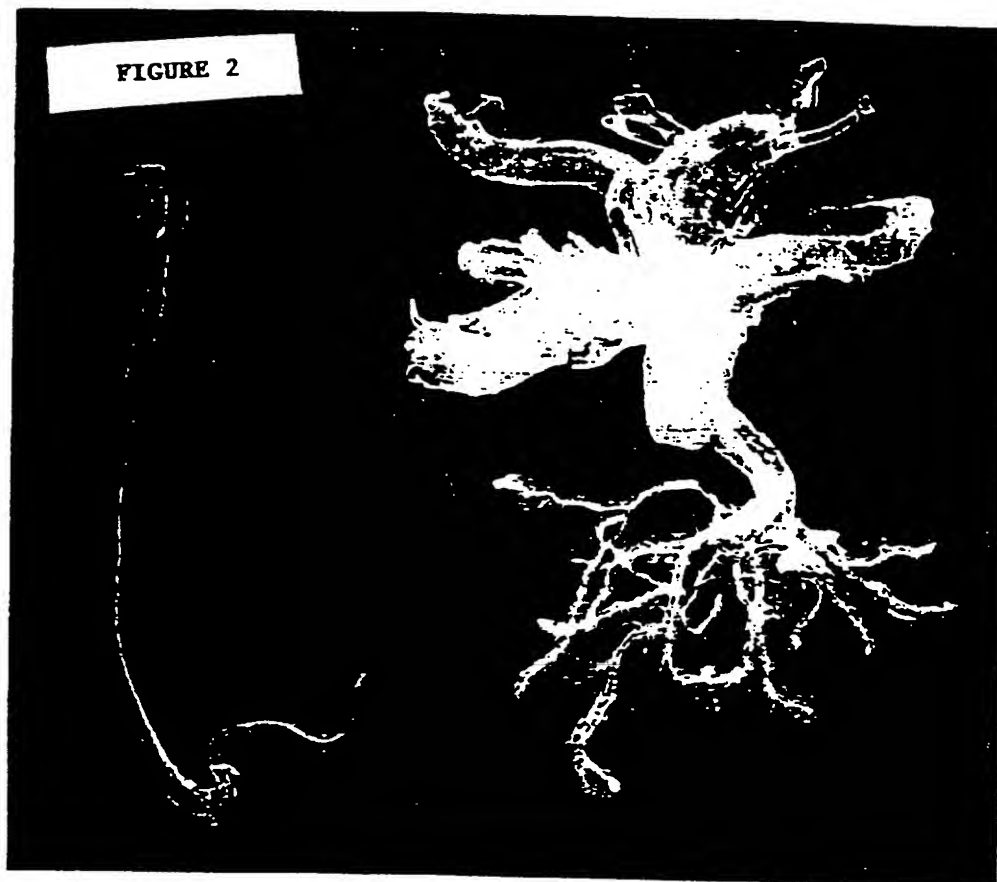
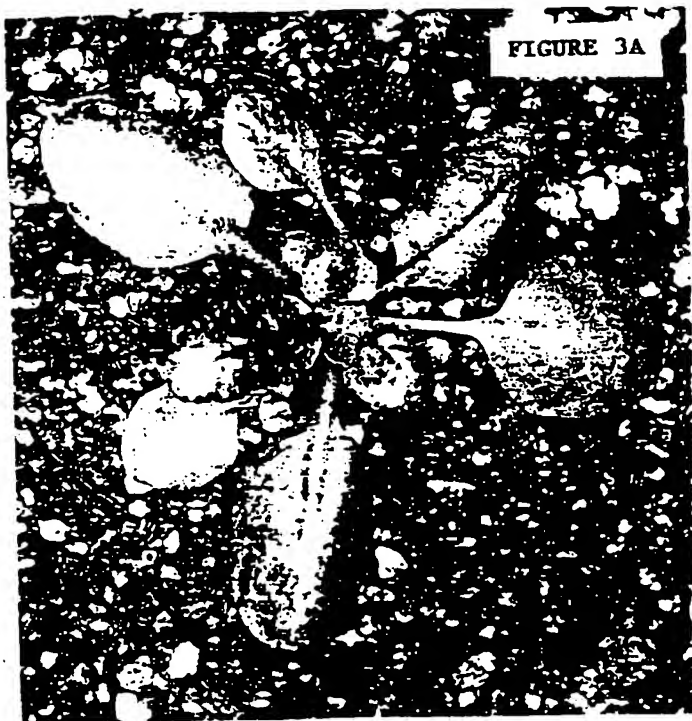


FIGURE 1B



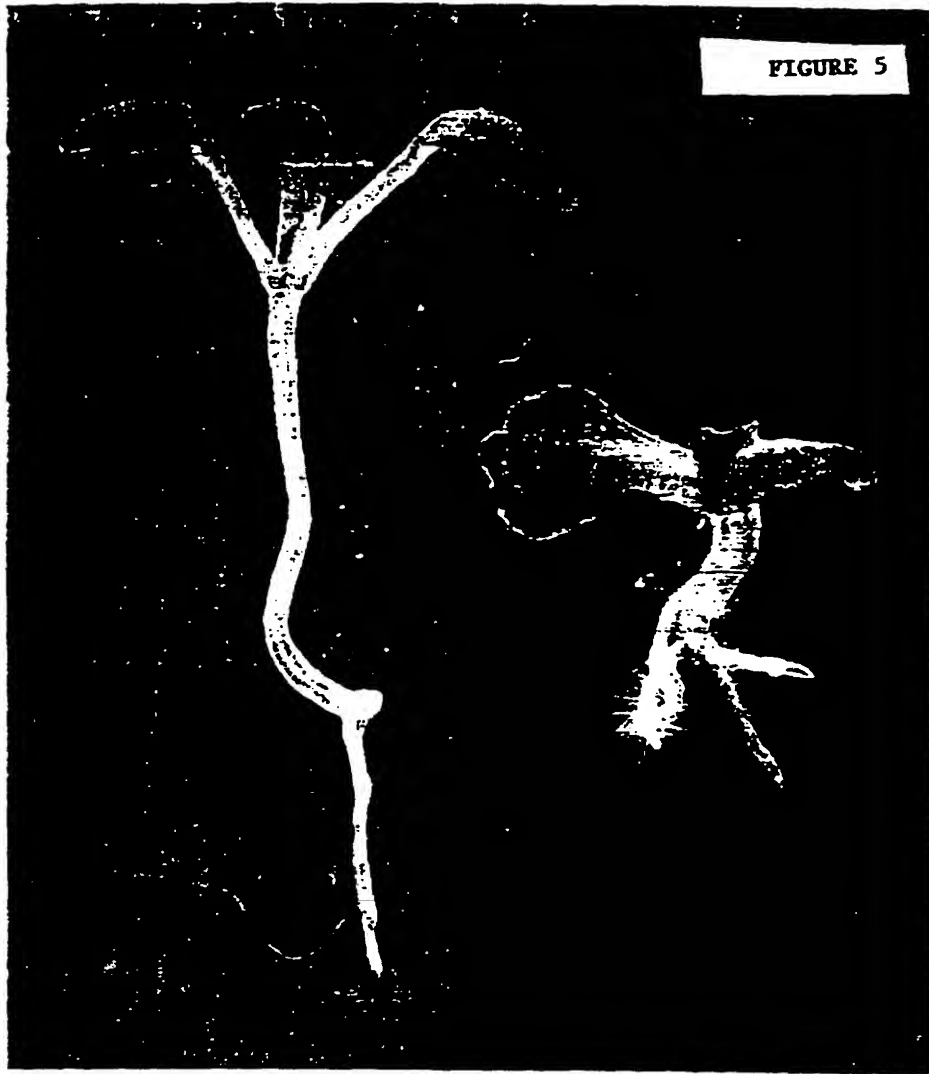




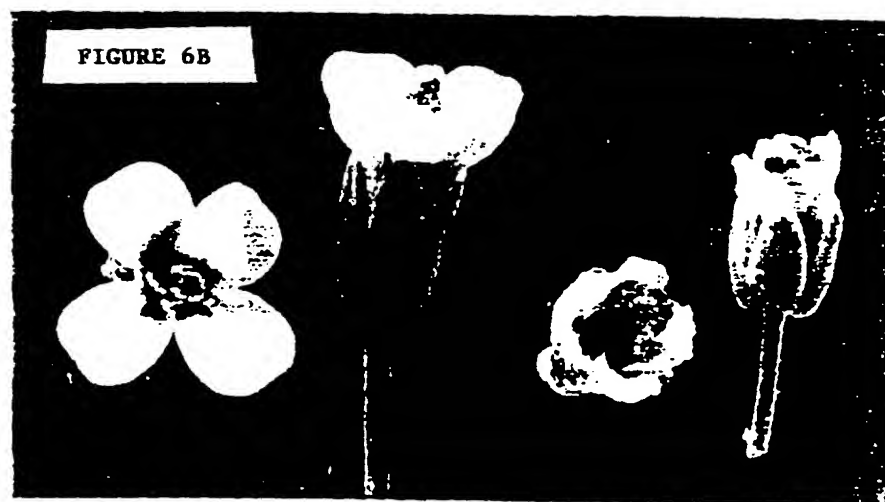
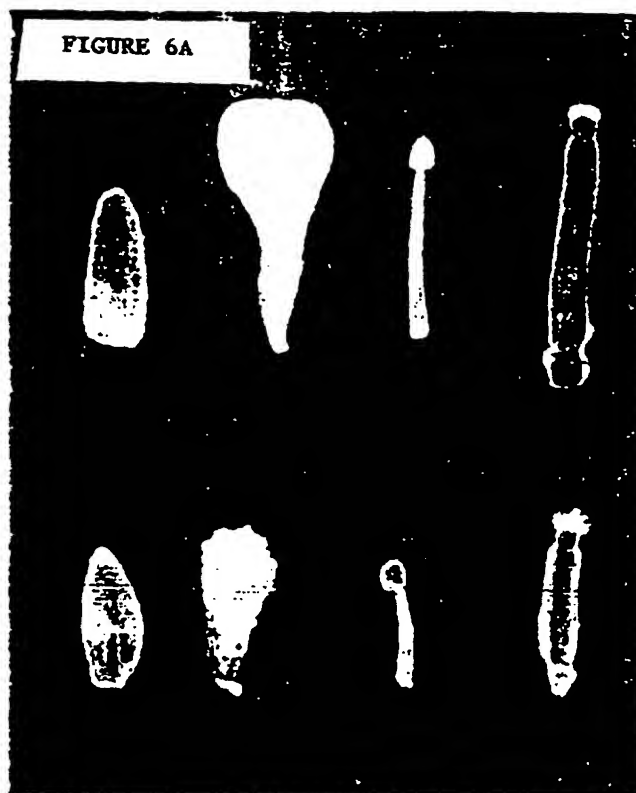
4/29

FIGURE 4





6/29



7/29

FIGURE 7A

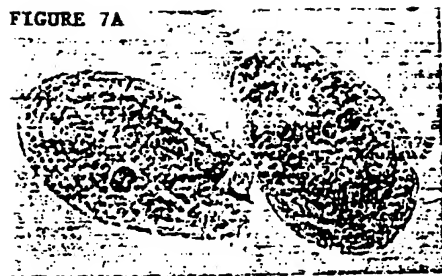


FIGURE 7D

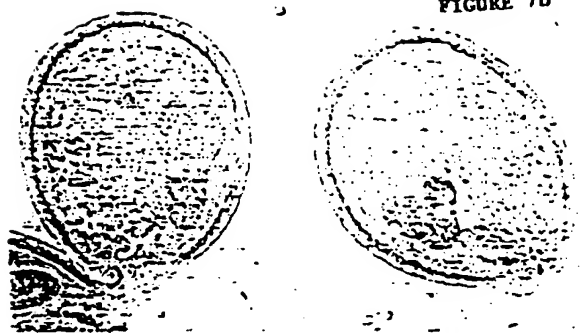


FIGURE 7B

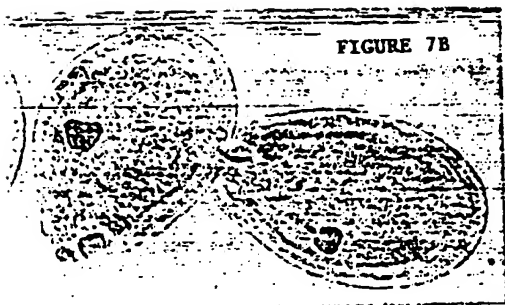


FIGURE 7E

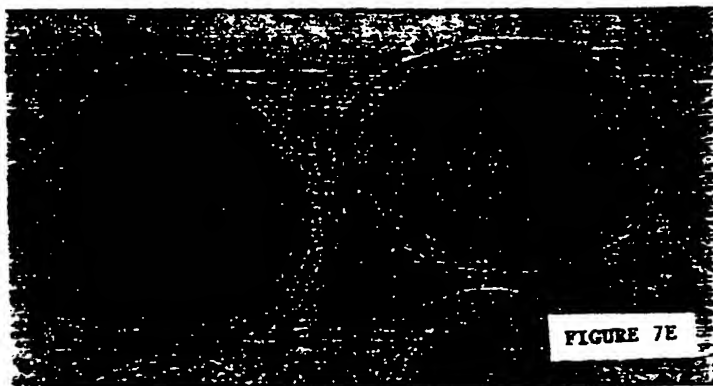


FIGURE 7C

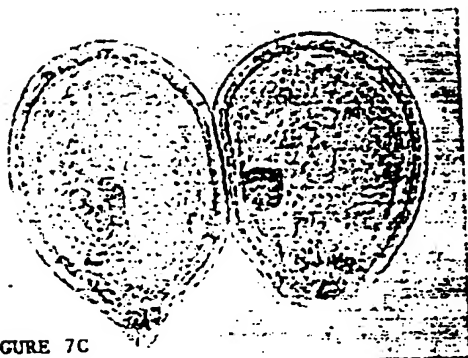
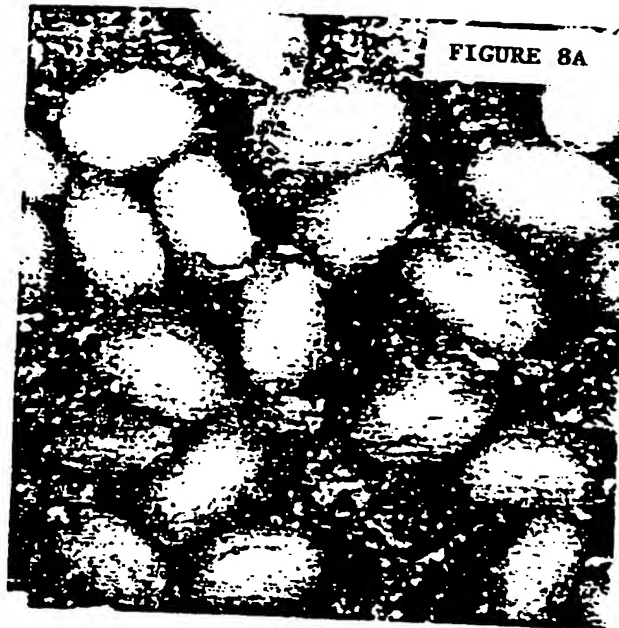


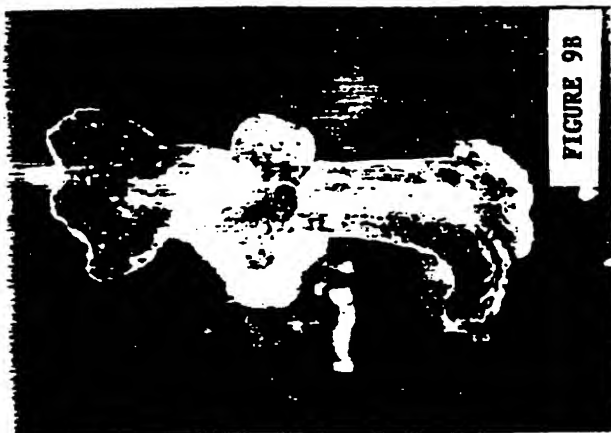
FIGURE 7F



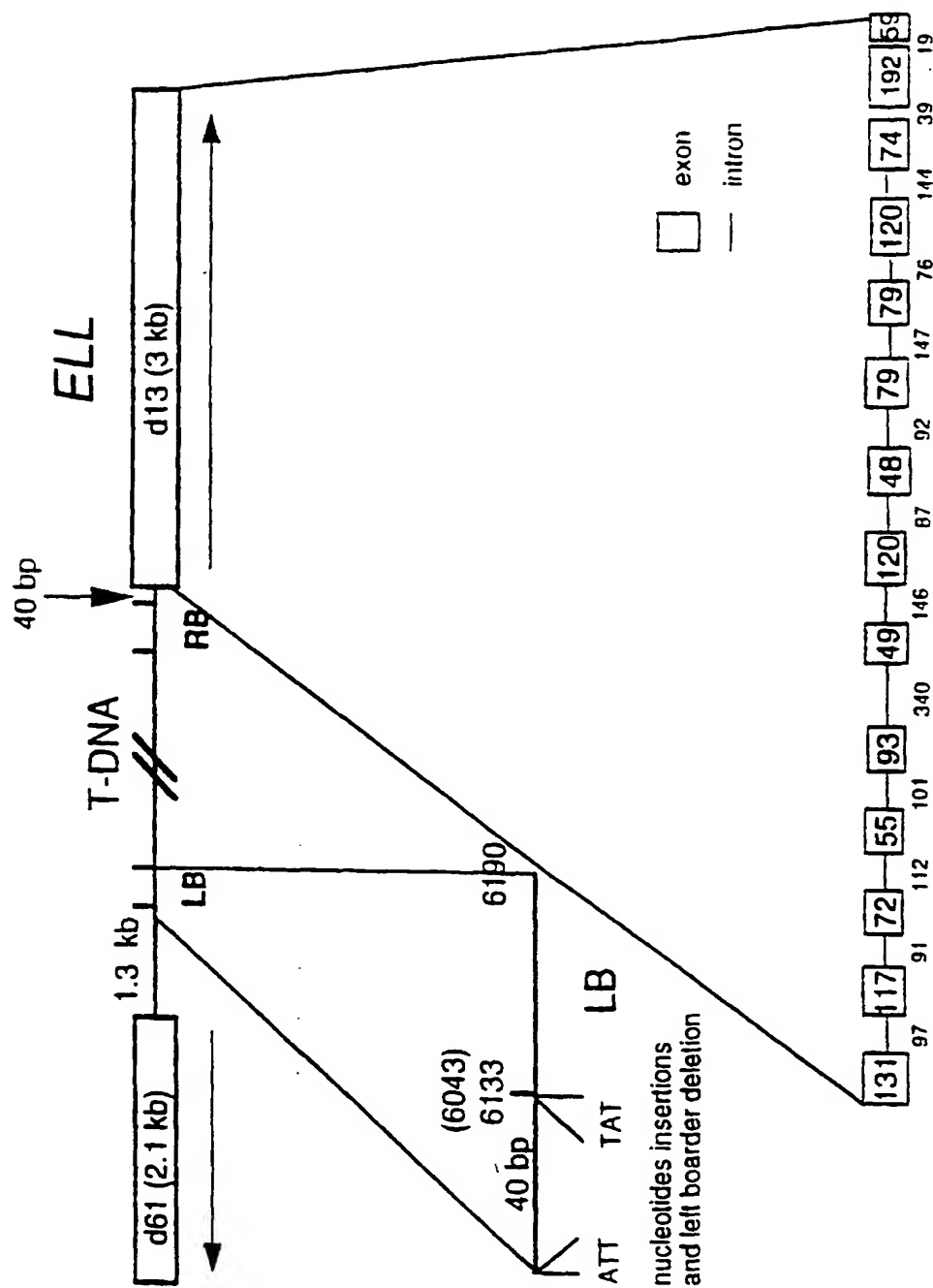
8/29



9/29



ELL encodes a putative C-14 sterol reductase



**deduced peptide sequence predicts
8-9 putative transmembrane domains**

FIGURE 10A

11/29

chr3

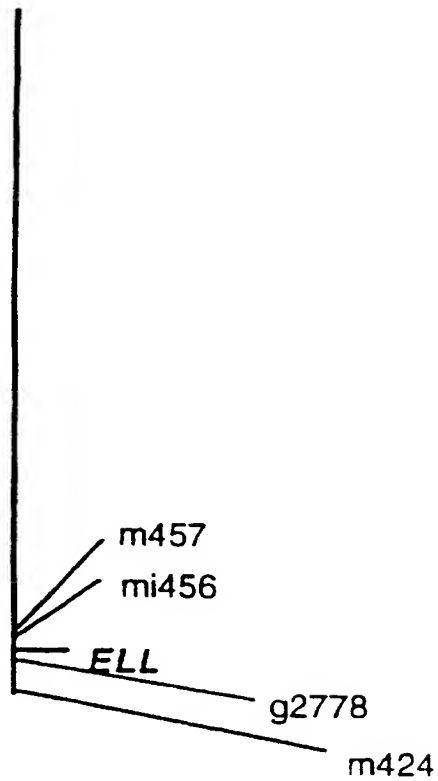


FIGURE 10B

FIGURE 11

Ach	MEENRKYADGE	VVMGCHWPGS	LYXKEVYVTSY	DDPGRLYTVR	41
Chicken	MPGGRKFADGE	VVFGRWPGS	LYYRVVYVLSH	DSISLYTVR	41
Ach	YKDGTELEALK	ESDIKPLQSSS	KORKSGSSSS	SPSRRRSGRS	51
Chicken	YKDGTELEALK	ENDIKPLQSSS	KORKSGSSSS	SPSRRRSGRS	51
Ach	RSRSHSPGHP	ARGRRSSSSH	SREHESOKKK	EEESLAP	61
Chicken	RSRSHSPGHP	PMASARRSASA	SHJADOKKAR	SESEVETLAP	61
Ach	PEPPEENNTNR	RYNGERDSSE	RNDTSSKLENE	QAKKLPDVEN	71
Chicken	ILKPPFCNS	RYNGERDSSE	RNDTSSKLENE	QAKKLPDVEN	71
Ach	ERVUCIOYSLR	SRREEKK	KEEHYAS	KKLPASART	81
Chicken	SYVUCIOYSLR	PRREEVK	DSKEEKYVA	KKLPASART	81
Ach	EKPSSTKREL	EPGGRFGTPM	LMFFLPAATVL	MLLD	91
Chicken	FVTPTIRAKDE	PFGGVVPGVFL	IMFGLPVFL	YLLLMCKOD	91
Ach	NDLGVL	FSLOSYV	VLVPYFV	YLAVAGEIL	101
Chicken	PSLHNFPPPL	PALESILWETK	VFGVFLWFF	FOALFYLLPI	101
Ach	GRVARGVLLS	DSQIRYRCN	GULALILLVA	ILGICAKLS	111
Chicken	GEVVEGLPLS	NPRKLOVRN	GFYAFLLTAA	AIGTLLYF	111
Ach	GRVVEGTPLI	UGRLKVRLLM	GFYAFYLTSA	VIGTSILPQGV	111
Ach	VSPLVYADRG	LELLSATFIF	CVLVTLLALYV	TGRSSSNKGS	121
Chicken	ELHYLYDHF	VQFAVSAAAF	SMALSILYLYI	RSLLKAPEE	121
Ach	FFHYVVSHF	LOFALAATVE	GVVLSVLYM	RSLLKAPRN	121
Ach	SLKPHVSGN	LVDHWWFGCO	LNPOFMSIDL	KEEFVHACH	131
Chicken	DLAPGCNSCZ	LVDHWWFGCO	LNPRIGSFDL	KYFCCLRPGL	131
Ach	DLSPASSGN	AVYDFHICRE	LNPRIGTFDL	KYFCCLRPGL	131
Ach	MUGLLINLSI	LAKSVODG	SLSSSMILY	QIFCALYILD	141
Chicken	IGWVVINILAM	LLAEMKIHNO	SMPSLSMILV	NSFOLLYVVD	141
Ach	IGWVVINILVM	LLAEMKIODR	AVPSILAMILV	NSFOLLYVVD	141
Ach	YFVHEEYMTS	TWDIIAERLG	FMLVFGDLLW	IPFTFSQGW	151
Chicken	ALWNEEAVLT	TMDITHDGGF	FMLAFGDLVW	VPFVYSLOAF	151
Ach	ALWNEEALLT	TMDIITHDGGF	FMLAFGDLVW	VPFVYSLOAF	151
Ach	WLLHNKVLT	VFAIVVNCLV	FLIGYMVFRG	ANKOKHIFKK	161
Chicken	YLVGHPIAIS	WPVAAATIL	NCIGYVIFRS	ANSQKNNFR	161
Ach	YLVSHNEVS	WPMASLIVL	KLCGYVIFRG	ANSOKNAFRK	161
Ach	NKXTPIWGKP	PVVVVGKX	LLVSGYWGIA	RHCNYLGDLM	171
Chicken	NPADPKLSYL	KVIPTATGKX	LLVSGWGFV	RHPNYLGDLI	171
Ach	NPSDPKLAHL	KTIHTSTGRN	LLVSGWGFV	RHPNYLGDLI	171
Ach	LALSFSLPCC	ISSPVYFYF	IYLLLLFNR	EARDEVRCAE	181
Chicken	MALAWSLPCC	FNHILPYFYF	IYFLLLVHR	EARDEHHCKK	181
Ach	MALAWSLPCC	FNHILPYFYF	IYFLLLVHR	EARDEHHCKK	181
Ach	KYKEIKASYL	PLVPWHIPY	VY	CKLYTSHLC	191
Chicken	KYGLAWERYC	QRVPYTH	HLLEHSTYLC	CKLYTSHLC	191
Ach	KYGLAWERYC	ORVPYRIFPY	HY	CKLYTSHLC	191
Ach					369
Chicken	TNSVCTYLGFK				H637
Ach					615

FIGURE 12

14/29

FIGURE 13A

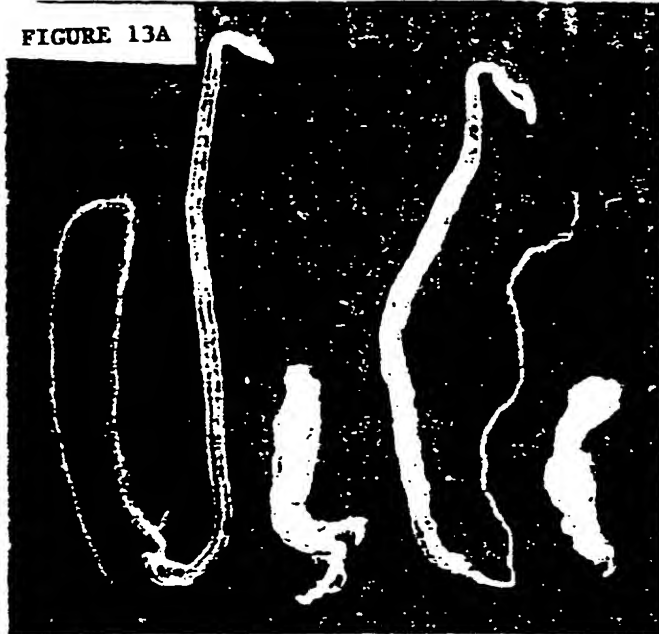


FIGURE 13B

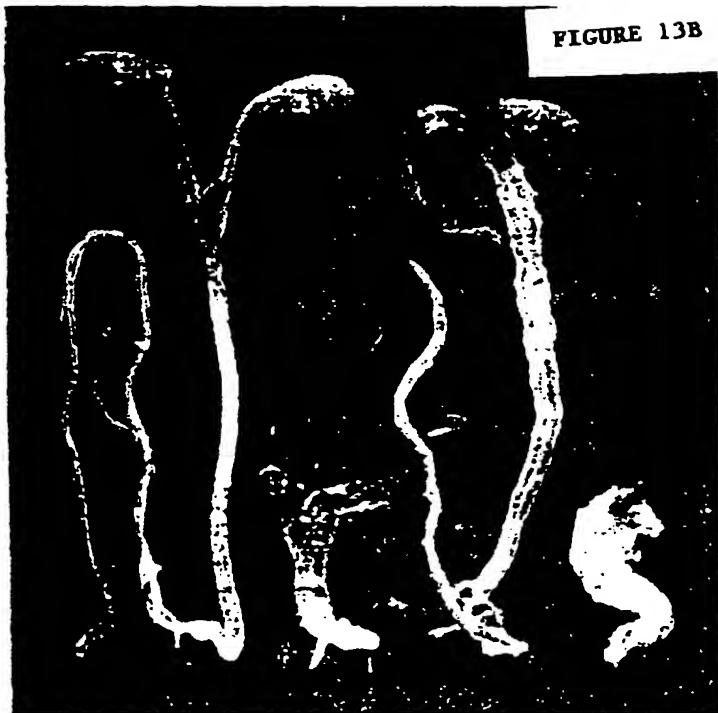


FIGURE 14 / SHEET 1

C

c

C

BNSDOCID: <WO_9748793A1_1_>

RNSDOCID: <WO 9748793A1_I_>

FIGURE 14 / SHEET 3

// // /

481 TACAGCTGAATCCTCAGTTTATGAGCATTGATCTCAAGTTTTCTTTGTCAGAGCCGGGA 540

ATGTCGACTTAGGAGTCAAATACTCGTAACTAGAGTTCAAAAAGAAACAGTCTCGGCCCT

c Q L N P Q F M S I D L K F F F V R A G M -

F
Cn
B B vuFT F C DN
c c i4os o v B rl
c c JHke k R c da
I I IIII I I I IV

//

541 TGATGGGATGGCTGCTTATCAATCTCTCTATTCTGGCAAAAAGTGTGCAGGATGGTTCCT 600

ACTACCCTACCGACGAATAGTTAGAGAGATAAGACCGTTTTTCACACGTCTCTACCAAGGA

c M G W L L I N L S I L A K S V Q D G S L -

B S
H H cM BB a
iF B TP iT eb gsDu
no s al nf 8o ltp3
fk g qe fi 3I IYnA
II I II II II IIII

/ / /

601 TGAGTCAGTCGATGATTCTTTACCAGATCTTCTGTGCGTTATATATATTGGACTACTTTG 660

ACTCAGTCAGCTACTAAGAAATGGTCTAGAAGACACGCAATATATATAACCTGATGAAAC

c S Q S M I L Y Q I F C A L Y I L D Y F V -

T
t
N
l
h
N
l
aB
If
Ia
II

N N T
l lM B
ab sBT M
Io asa n
II Ali l
II III I

A
p lC B
5 wv s
0 2i m
9 6R F
I II I

C B M C C
j B v v
e s w i i
P m o J J
I I I I I

/

661 TTCATGAAGAAATACATGACCTCTACGTGGGACATAATTGCAGAGAGACTAGGCTTCATGC 720

AAGTACTTcTTATGTACTGGAGATGCACCCTGTATTAACGTCTCTCTGATcCGAAGTACG

c H E E Y M T S T W D I I A E R L G F M L -

S
C BB a H C C C
M j gsDu iT j B M v v
s e ltp3 nf e s w i i
l P IYnA fi P m o J J
I I IIII II I I I I I

/ / /

721 TAGTGTGGAGATCTCCTGTGGATTCTTTTCACTTTTAGCATTACAGGGCTGGTGGCTTT 780

ATCACAAACCTCTAGAGGACACCTAAGGAAAGTGAAAATCGTAAGTCCCGACCACCGAAA

c V F G D L L W I P F T F S I Q G W W L L -

T

FIGURE 14 / SHEET 4

C T s
 v s A p M
 i p l M5 Bb
 R 4 w u0 bo
 I C N n9 sI
 I I I II II
 / /

TGCACAACAAAGTAgAACTAACAGTTCCTGCGATTGTAGTCAATTGCCTTGTCTTCTTGA
 781 ----- 840
 ACGTGTGTTTCATcTTGATTGTCAAGGACGCTAACATCAGTTAACGGAACAGAAGAACT

c H N K V E L T V P A I V V N C L V F L I -

N
 l
 R aM T C B
 s In a Av s lM
 a Il q uJ R ls
 I II I II I II
 /

TAGGGTACATGGTTTTTCGAgGAgCTAACAAACAAAACATATCTTTAAGAAGAACCCAA
 841 ----- 900
 ATCCCATGTACCAAAAAGCTcCTcGATTGTTTGTGTTTTGTATAGAAATTCTTCTTGGGTT

c G Y M V F R G A N K Q K H I F K K N P K -

M
 a
 b BM B C a v B jMs e C B v
 o ps s j c i s enp I j s i
 I ml l e 8 J r PlR I e r J
 I II I I I I I III I I I I
 //

AAACACCAATATGGGGCAAGCCTCCAGTGGTAGTTGGTGGAAAGTTACTGGTTTCAGGCT
 901 ----- 960
 TTTGTGGTTATACCCCGTTCGGGAGGTCACCATCAACCACCTTTCAATGACCAAAGTCCGA

c T P I W G K P P V V V G G K L L V S G Y -

T
 s
 C p C T s
 j 5 v s pT B S l E
 e 0 i p 5s sSf l c HC
 P 9 R 4 0p ata l 5 nc
 I I I C 9R JyN I 7 48
 I I I I II III I I II

ATTGGGGAATTGCAAGGCACTGTAAATTACCTTGGCGACTTGATGCTTGCTCTGTCTTCA
 961 ----- 1020
 TAACCCCTTAACGTTCCGTGACATTAATGGAACCGCTGAACTACGAACGAGACAGGAAGT

c W G I A R H C N Y L G D L M L A L S F S -

N
 l
 a X B N
 I m sM l
 I n as a
 I I Wp I
 I I II v

FIGURE 14 / SHEET 5

1021 GTTTGCCATGTGGAATAAGTTCTCCGGTTCCATATTTCTACCCGATATACCTTCTGATAC 1080
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CAAACGGTACACCTTATTCAAGAGGCCAAGGTATAAAGATGGGCTATATGGAAGACTATG
C L P C G I S S P V P Y F Y P I Y L L I L -
A
1B M C
E ws M bT v RT B
a 2m n oa i sa s
r 6B l Iq R at g
I II I II I II I
/
1081 TATTGATATGGAGAGAACGAAGAGACGAGGTTTCGATGTGCAGAGAAGTACAAGGAGATAT 1140
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ATAACTATACCTCTCTTGCTTCTCTGCTCCAAGCTACACGTCTCTTCATGTTTCCTCTATA
C L I W R E R R D E V R C A E K Y K E I W -
E
B BcS B
s D soc
m d aRr p
F e JIF m
I I III I
/
1141 GGGCAGAGTATCTTAGACTTGTCCCTGGAGAATACTTCCTTATGTTTATTAGATGTGCC 1200
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CCCGTCTCATAGAATCTGAACAGGGGACCTCTTATGAAGGAATACAAATAATCTACACGG
C A E Y L R L V P W R I L P Y V Y *
N
C H1 H M
v R iaT.X H iT n
i c nIf m p nf l
J a fIi n h fi I
I I III I I II I
//
1201 AAGAGCCAAGTCATGAATCCTTTTCAGATTCACCTCTTGTGTCTTATTTTTTCCATAATC 1260
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
TTCTCGGTTTCAGTACTTaGGAAAGTCTAAGTGGAGAACAACAGAATAAAAAAGGTATTAG
C
T E
s c
B p RT o
sT5 sa 5
ra0 at 7
Dq9 II I
III
/
1261 TTGTTTTATTTTAGCAATGCTCGAATTGAACTTTGTAGTACACTTTTGAAAAATAACTT 1320
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
AACAAAATAAAATCGTTACGAGCTTAACTTTGAACATCATGTGAAAACCTTTTATTGAA
C
FF HM
C BnnGas
BAEEvAsuudepNTCT
M C

20/29

FIGURE 14 / SHEET 6

```

      s          j          scaaici44iIAoja
      e          e          liegJiEHHiIltueu
      I          I          IIIIIIIIIIIIIII
                               /  / / / / / / /
1321 CAGTCCTTAAAAAAAAAAAAACCTAANTTACTCCCNCCTGGGCGGCCGCTGGTTTTATAT
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      GTCAGGAATTTTTTTTTTTTTTTGGATTNAATGAGGGNGACCCGCCGCGGACCAAAATATA
                                     1380

```

```

      T          T          S
      s          s          a
      p          p          u D
      5 M        5          3 p
      0 s        0          A n
      9 e        9          I I
      I I        I
1381 TTGTTGTAAAAATTAAANAATTACTNCCTTGANGATCTGTAAAAA
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      AACAAACATTTTAAATTTNTTAATGANGGAAC TNCTAGACATTTTTTTTTT
                                     1429

```

Enzymes that do cut:

AcuI	AluI	AlwI	Alw26I	AlwNI	ApoI	AvaI	BbsI
BbvI	BccI	Bce83I	BcefI	BfaI	BfiI	BglII	BplI
BpmI	BsaAI	BsaBI	BsaHI	BsaJI	BsaWI	BseRI	BsgI
BsiEI	BslI	BsmI	BsmBI	BsmFI	BsrI	BsrDI	BsrGI
BstYI	Cac8I	CjeI	CjePI	CviJI	CviRI	DdeI	DpnI
DrdII	EaeI	EagI	EarI	Eco57I	EcoRII	Fnu4HI	FokI
GdiII	HaeIII	HgaI	HgiEII	Hin4I	HinfI	HphI	MaeIII
MboII	MnlI	MseI	MslI	MspI	MspAI	MunI	MwoI
NciI	NlaIII	NlaIV	NotI	NspV	PleI	PvuII	RcaI
RsaI	Sau3AI	ScrFI	SfaNI	SmaI	SspI	StyI	TaiI
TaqI	TatI	TauI	TfiI	ThaI	TseI	Tsp4SI	Tsp4CI
Tsp509I	TspRI	Tth111II	XmnI				

Enzymes that do not cut:

AatII	AccI	AceIII	AflII	AflIII	AhdI	ApaI	ApaBI
ApaLI	AscI	AvaII	AvrII	BaeI	BamHI	BanI	BanII
BcgI	BcgI	BclI	BglI	BmgI	Bpu10I	Bpu1102I	BsaI
BsaXI	BsbI	BscGI	BsiHKAII	Bsp24I	Bsp24I	Bsp1286I	BspEI
BspGI	BspLUIII	BspMI	BsrBI	BsrFI	BssHII	BssSI	Bst1107I
BstEII	BstXI	Bsu36I	ClaI	DraI	DraIII	DrdI	DsaI
EciI	Eco47III	EcoNI	EcoO109I	EcoRI	EcoRV	FauI	FseI
FspI	HaeI	HaeII	HhaI	HincII	HindIII	HpaI	KpnI
MluI	MmeI	MscI	NarI	NcoI	NdeI	NgoAIV	NheI
NruI	NsiI	NspI	PacI	Pfl1108I	PflMI	PinAI	PmeI
PmlI	PpuMI	PshAI	Psp1406I	PstI	PvuI	RleAI	RsrII
SacII	SalI	SanDI	SapI	Sau96I	ScaI	SexAI	SfcI
SfiI	SgfI	SgrAI	SimI	SnaBI	SpeI	SphI	SrfI
Sse8387I	Sse8647I	SstI	StuI	SunI	Swai	TaqII	TaqII
Tth111I	UbaDI	VspI	XbaI	XcmI	XhoI		

FIGURE 15 / SHEET 1

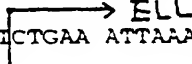
E11.Tmp6 Length: 6588 May 3, 1996 10:36 Type: N Check: 7899 ..

```
1  TNTTGAAGGN TNAAGAAAAA NTANGGTAAG CTGGGNAGGA CAAGANTTCT
51  TGTNACCACA ACACAACAAC GCCATGAACC NATCGGTTTC TTNTGTTTNG
101 AGATCACCTT TCTTGAGTTG GTGGTTTCTG AGNTCAAGNT CCTTGTTGAC
151 TCAGTGAAGT CCAGATGCAG CNTCAAAACT TTTGTCCTGT AGACNTAGCA
201 AGAGTAACAG CACCAACCAA ATCGCTATCC GATGTAATCA AAACCTTATC
251 ACNTTCATCG TCCTCATATA TAATCTGAGG CCGTTGTTCC ACATTGTTAT
301 TGTCGCTGCC AATTCTTTGC ATCACAATAC CCATCAGCTN TTCGAGGTTT
351 TCAGCTCCAG AAGTAAACCG ATGTACACGG CCCTTAAGGT CTTCAAATTT
401 GAACGAAAAC GAATTCCCTA GTCCTAGAGA TGGGTAAGAA CTGAGCTTCC
451 CTATATCTGA ATGATGCATC ATTGCCGACA TTTCACTTTG AGTGTCAGAA
501 TCATCAGGTG GCTCTAACGC AAGAGCTGAA TCCCAAAATT TCTGCATCAT
551 CGTGTGTTGCC ATATCATTTA CAGCTCCAGA ACTGTTCTCC ACCATTGAAA
601 TAGCTGCGTG AGTAATCTGA AGAACGTCTA CACAAGCTGC AGCTGATCCA
651 TCTTTATCTA TAATTGGAAG ATGTAGAAAC TTTCCATCAT GCATTSTATG
701 CAATGCATCC AGAATCGTTG TCTCTAGCGA TGCACATTCA GGATTCCGTG
751 TCATTACCTT CTCGACAAGA GTCAATTCAG GAGATAAATT TTGTGCCACC
801 ACTCGCATCA GAATGTCCTT TGAAGTCAAG ATTCCACTGA TTTTGTTCCC
851 CCGTGAAAT GATTACAGAG TTAACCCGCA AATCCCTCAT CCTTTTCGCA
901 GCAACTGAAA CAGGATCTGA TGSTGCTACA AGTGCAACCT TCGATGTCTG
951 TAATAATCGT GACAAGGCGG GTTAAACATT NTNTCCTTCA AGGTTTCAAT
1001 GAAAGCATAC GSTGCAGAA ATCCGCTTCC CCATTGTTTN TCCACACCTT
1051 CCACTGCAGC AGCTAAAGCA CTACCTTGCT CTGCAGTTTC TCCATCCTAG
1101 AAATAGCATC ATACAAACAC TTTGTAATAT CCAACAAAGC AATGACTTCA
```

FIGURE 15 / SHEET 2

1151 CCATTCTCCA CAACAGGCAA GTGTCTAAAC TTCCCTTGAA CCATCTTCTG
1201 AAGAGCCTCA AGCGCCAACG AATCAGAAGT AACAAAAATA GGATTCCCTAG
1251 TCATAACCTT AGAGACCAAA GTTTGATCCG GTCTCAACCC TTCAGCAATC
1301 ACTCTTGTAG CTACATCTTT ATCAGTAACA ATCCCGGAAA GAAGCGCACT
1351 TGAATCAGTC AACAAACAAG CATCAACACG CCTAGCAGCC ATTNTTCGAC
1401 AAGCATCGAA AACAGTAGTT CCTTCNAGGA ATAGTAAGAG CTTTCGATAA
1451 CCTAAGCTTC TTCGCTGTCT CTCTCCATTA GAAGGAGCTT GAGATTGAGG
1501 TTCAGGAGGA GGTGAATTGG GTTTTGAGGT GTTCCCAANTA ACACTTCCAT
1551 TCTCTGATTG TACTGGTTTC TTAGAAGGTG GTGGTCCTCN CCGTACAGTA
1601 GAATTGCTTC TCCTCCCTGA TGTTGAAGAA GGACCCGTCG CTTGAGTACT
1651 CATATTGGGT CAATCTAGGG TTTACTTAGA TCCTAAATCC GTCANAAATG
1701 ATTCTTTTAG ATATCAAACG CGTCTCTGCA AATGAAAAAT TCAACCTTTA
1751 ATTCACAAAC TATTGAAATT TCATCTAAAG CACGAATCTG AATAAAACCC
1801 AATTCACAAT AAAGACGATT TGCTCTGAGA ATACGATGCA ACATACACGA
1851 AAAGGATTCTG AATTTAACGG ACGAGGGAAA TGAAACAACG TGAAACCCCTA
1901 AGGATTTGAG CAGAAGTTAT GTGGGAAGAT TGGGNATTTA GGGTTTACCT
1951 TCTTCTTTCT TCNTCAAGGT CTCTCTCTCG AGCACTTTCTG TTNCCCCAAA
2001 AACNAACGGC TCTTAACAAT TGAGTTAANC CANTTATCGA GTTTTCATTG
2051 GNTGTTCCCTG TTTCCGCGTG TGTGGTGGNT CNCCACCTCC TTTCTTATAA
2101 TCNACGACTA AAAATGTTAA ANATAANACT AANATTTCTT TCTANAAAAA
2151 TCGTAAAANC CAAATGTTTT TTTTTTCTG ATAAATGTCT ATAAATCACC
2201 CTTTCTTTTT AAATAATGAA ATTTGATGAC ATTTATCTCT TGTATCTAGN
2251 AGAGTTAATG GCTAACATAA ANACCAAAAA AAATTAATTC NAATAAATAT
2301 GATTTGTGTG GGTACATGG AAAAAATGTC AAATAATAAA NCAAAAAAAA
2351 ATTGATATAGA TGCAGTGCAA GTTGTTTCTG GTCAACTTGC CGTCGAGCCT
2401 CACAACGTTT TGTTACAAGT GGACTCGCAT GTAATTCCTT CTTTTAATAA
2451 CTTACCAGTT ACACCATCCA ACATGTGATT TGACAGAAAA ATATTTTACT
2501 GAAATGTGAT CGGTGCAGAT TTTTCTATGT ACGTTTAAGC CTTTAAGGTA
2551 GACGTTTAAT CCNAAAATAT CCCTGAATAA CAACACCGAT TAATGGAACC
2601 AAGTAGATAC CTCCTCCGTT TGGATGGCTC AAATGCAACC ATGATGCAAG
2651 CTTTTCGGAT TGACCCAAAG TGAGAGAACT AGATCGAGAT GGATTATTCC

FIGURE 15 / SHEET 3

2701 GAACCATTAC GGCACCCTTA TATAATGGCA GCATCTTAAT AGTAAACAAA
2751 AGCTTTAGCC TTAGGTTTFA GCTTCCTTCA CTCTTTGCAT ACATTGTGAA
2801 TCTGCGGTTT TAGATGGACC ATAGTGGAAG AAGGCTTTCA TCAATAACTC
2851 GTGGACTTGA TCAATGGTAG AAAAGANAAT ACATAGTATG GAAAACTAGA
2901 TATTTGATAT ATTTGGTTCA AACTCTTATC CCGTGTTGAG GTGATATACA
2951 CATGAAGACA TAACAATCGC ATAGCCGAGA AACTAGTATT CATTAACCTT
3001 TTTCTCTAAA GAGATTGTCC TATCAATCTA AATTTTAGAT GTTAAAAAAA
3051 AAATggttaag gttaaacagg ccgctagggtt ggtttttacga tgatgtaaaa
3101 agtagccatc ttaaaataac agtcggtttgc gagactggcc aggccatccc
3151 atggggcata ggctcgctca agttgTGCTT GGCAGAATTT AGTAACTTGG
3201 GGTTTTGTTA TCAACAATCA ATAGTTTAAG GCTTTACCTG CAAGAAATGA
3251 AGAGTTTAAG GGTTCCTTTT GGTATTCCCG ATTACACAAA GTGAGCTAGC
3301 TCATCAGAGT CCACGAGCTT CCCACTAAAA AATTGAAAAT TGTTCCTTCT
3351 GTCATCTGAA  ELL ATTAAACAAA GCGAGAAAAG GCGATACAAA CGATTTGAA
3401 TGCTTCATCT TCTCCTTGA AAATCCTTCT TCTGCTTAAT GCTGCTAGAT
3451 ATGGATCTCG GTGTTCTTCT TCCATCATTG CAATCTGTGA GCTGTCTCTT
3501 TAGCTTTTGA CTGTTGCAAT TGTTATTGTG AAATTTTGTG TCGCTTTTGG
3551 ATCAGCTTTT GTTAAATTCG TTCCGAGATT TTAGGTTTAT GTGCTGGTGT
3601 TTTACTTCGT TTAATTGGGN CGNTGGCGGA GAAATTCTCC CCGGGAAAAGT
3651 TATTCGCGGC GTCTTTTAT CAGATGGCTC TCAACTTCGT TACCGATGCA
3701 ATGGTATATT TGATTTGATT TACTCTCTCT ACAATTCTCTG AGAGTCTGTG
3751 AGCTCGAAAG TTCATTTCCA TTAGTTTGGT TAATTCAATT TCAGGTCTAT
3801 TGGCACTAAT ATTGTTGGTA GCTATTTNGG GAATCTGTGC AAAACTTGGC
3851 ATTGTATCAC CTCTTGTAAG TGTAGTTACA AGATTTGAT TGTATTTCTA
3901 TGAATCCGAA TGCTATATGC TATATGAATC CGATTGCAAT TGCTTTCTCA
3951 CACTCATTCG ACTGAGATGT TTGGTAGGTG GTTGCGGATA GAGGACTTGA
4001 GTTACTCTCA GCTACTTMINA TTTCTTGTGT TTGGGGAAGA TGATCAATCC
4051 TTAGTCCGNN GTCTTGGATT TTAGNTGNCT TACCATCAGA TTNGCTTTGG
4101 GTGGTGTGAT TTGTAATCTC CATGATATCT CTTAATATTC TCAGGTGACA
4151 TTASCATTGT ATGTTACTGG GCGAAGTTCC TCGAATAAGG GTTCTTCCCT
4201 AAAGCCTCAT GTCTCAGSAA ATCTTGTACA TGACTGGTAC TAACATAATA

24/29

FIGURE 15 / SHEET 4

4251 CAATTGTAGA TCTGATACTT TCTTGTTACA CAAAATGTTG TTTAAAAGTTA
4301 TATATTTTGA CTCCTGCAAG AGCAAACTA AGAAATAATC TGGTACTATA
4351 TAGAGTTTGA AACACTGAAT TGGACAAGAT GATTCTATAG AACTTCGTAG
4401 AGTGTTGAGT AATTTCTCCT AGAACGTTG TAGCTTCCTC TTTTTCCTT
4451 TTAACCGCAG TGACTTTAGC TTTTGGAAC TTTCTACTGA AACTAGAAGT
4501 TCTGGTTTTG TCTTTCACCT ATCTCTTCCA AACAACTGCT TCAATTTTTT
4551 CTCATATTGT TTGTTTCATG TGATAGGTGG TTTGGAATAC AGCTGAATCC
4601 TCAGTTTATG AGCATTGATC TCAAGTAATC CATTTTTCTG TTTTTCCTC
4651 TATTTGTCAG CCAAGGCTAC ATCATTGCTT CAGTTTGTTC CGTACTCAAT
4701 CGAGTGGCAG TTTAATAATG TAATCAGCAG TTATGCATGG TTATGATGAA
4751 TGGGAGTTAT TCCTTGTTGA GGTTTTTCTT TGTCAGAGCC GGGATGATGG
4801 GATGGCTGCT TATCAATCTC TCTATTCTGG CAAAAAGTGT GCAGGATGGT
4851 TCCTTGAGTC AGTCGATGAT CTTTACCAGA TCTTCTGTGC GGTAAATTTG
4901 GTTTTTACTT ACAAATCTTG CTCTTGAAN TCTGATCATC TGTGTTTTGT
4951 TAGTTTGTAT TAGTTTTATA ATTGCAGTTA TATATATTGG ATACTTTGTT
5001 CATGAAGAAT ACATGACCTC TACGTAAGTT CATGGCGTGT TAAGGAAACA
5051 CATTTGTCTT ACCAAAAAAT GACCATTGTC ATTATTACAT CTACTTTGAT
5101 TTTACTCTTT TCAGGTGGGA CATAATTGCA GAGAGACTAG GCTTCATGCT
5151 AGTCTTTGGA GATCTCCTGT GGATTCCTTT CACTTTTAGC ATTCAGGCAT
5201 GTAACGTGTA GCCTGAACAC AAACAAGATA TTAATTTATC TTATTGACAG
5251 TATCTTCTTG GCATGTTACA GTTATTCTCG GAAACAATAT TGTCTAGAA
5301 TGCTTGATCA CTCTGTGACT GAATTGTCTT CTCTCTGGTA CAGGGCTGGT
5351 GGCTTTTGCA CAACAAAGTA GAACTAACAA TTCTTGGAT TGTAGTCAAT
5401 TGCCCTGTCT TCTTGATAGG GTAAGTTCTG AGACATGGGG TTATTTTCCA
5451 TTCTTACATA TCTACACTAA GAAACCCACT ATTTCTTCTT TGGCAGGTAC
5501 ATGGTTTTTTC GAGGAGCTAA CAAACAAAA CATATCTTTA AGAAGAACCC
5551 AAAAACACCA ATATGGGGCA AGCCTCCAGT GGTAGTTGGT GGAAAGTTAC
5601 TSGTTTTGAG CTATTGGTAT GTTATATTTA TCTTCTCTTG TTTCTTTGCT
5651 TSGTTTTGCC ATCTCTGTGT TSGATTGTTT ATCATGCTGG GAATAAAGAG
5701 TTGAAAAGTTT CGCAATGACA CATTTCCGAT AACTTAGGTG CTGTTTTGTA
5751 TATATGACAG GGGAAATTGCA AGGCACTGTA ATTACCTTGG CGACTTGATG

FIGURE 15 / SHEET 5

5801 CTTGCTCTGT CCTTCAGTTT GCCATGTGGA ATAAGGTACT CCTNCTGCTT
5851 GAGTTCACCTT ACAGCTACCA AAATCATGTA GAAACTAATA CCAATATCNA
5901 AACGTTTCGAA GTTGATTTGG CTGACTTAAA GATATTGATC TCTAACCATC
5951 ATTTGAAAAG TCTAAAGCTT TCAAGTTCAT TTCCCAAAGC TGTTTTTTATG
6001 ATATTTTCGTC TNGTGTATTC TCAGTTCTCC GGTTCCATAT TTCTACCCGA
6051 TATACCTGCT GATACTATTG ATATGGAGAG AACGAAGAGA CGAAGTTCGA
6101 TGTGCAGAGA AGTACNAGGA GATATGGGCA GAGTATCTTA GACTTGTCCTC
6151 CTGGAGAATA CTTCCCTTATG TTTATTAGAT GTGCCAAGAG CCAATTCATG
6201 AATCCTTTCA GATTCATCCT CTTGTGTCTT ATTTTTTTCAT TAAATGTGAC
6251 NTGAAATGAT CCCATTATNG CCTNTTATCA ATGCTTGATT GAAACTTTGT
6301 AGTACACGTT TGAGAATTAC ^{ELL} ← TTCAGTCCTT | GTTATTATTT TAGCATGGAT
6351 ATCAACATTT TCGGATTTAT TTNTNGGGTT ATTTTAAAAC CNNAGATTAC
6401 CNAANAAAAC CATTGTTTGA NGTANGATAA TATGGACTTT TTA CTGAAAA
6451 AAAATNCTAN TAGGGGAACA AATNGAAGTT GAATATGGCT GAATNTTTTT
6501 ATGGANAAAA TGGAAACTTT TCCCACTTTG AAATGACAAT NCAAGTTTGG
6551 TGGACNACTT AATCACTGGA AACGTTAATG GCCAACCN

26/29

FIGURE 16
SHEET 1

```
1 CTGAAATTAAACAAAGCGAGAAAAGGCGATACAAACGATTTCGAATGCTT 50
|||||
3356 CTGAAATTAAACAAAGCGAGAAAAGGCGATACAAACGATTTCGAATGCTT 3405

51 CATCTTCTCCTTTGAAAATCCTTCTTCTGCTTAATGCTGCTAGATATGGA 100
|||||
3406 CATCTTCTCCTTTGAAAATCCTTCTTCTGCTTAATGCTGCTAGATATGGA 3455

101 TCTCGGTGTTCTTCTTCCATCATTGCAATCT..... 131
|||||
3456 TCTCGGTGTTCTTCTTCCATCATTGCAATCTGTGAGCTGTCTCTTTAGCT 3505

.
.

132 .....GTTTATGTGCTGGTGTTTTAC 152
|||||
3556 CTTTGTGTTAAATTCGTTCCGAGATTTTAGGTTTATGTGCTGGTGTTTTAC 3605

153 TTCGTTTACTT.GGCCGTTGCCGAGAAATTCTCCCCGGGAAAGTTATTC 201
|||||
3606 TTCGTTTACTTGGGNCGNTGGCGGAGAAATTCTCCCCGGGAAAGTTATTC 3655

202 GCGGCGTCCTTTTATCAGATGGCTCTCAACTTCGTTACCGATGCAAT... 248
|||||
3656 GCGGCGTCCTTTTATCAGATGGCTCTCAACTTCGTTACCGATGCAATGGT 3705

.
.

249 .....GGTCTATTGGCA 260
```

FIGURE 16 / SHEET 2

3756 GAAAGTTCATTTCATTAGTTTGGTTAATTCAATTCAGGTCTATTGGCA 3805
261 CTAATATTGTTGGTAGCTATTTTGGGAATCTGTGCAAACTTGGCATTGT 310
3806 CTAATATTGTTGGTAGCTATTTTNGGGAATCTGTGCAAACTTGGCATTGT 3855
311 ATCACCTCTT..... 320
3856 ATCACCTCTTCTAAGTGTAGTTACAAGATTTTCGATTGTATTTCTATGAAT 3905
321GTGGTTGCGGATAGAgGACTTGAgTTAC 348
3956 ATTCCACTGAGATGTTTGGTAGGTGGTTGCGGATAGAGGACTTGAGTTAC 4005
349 TCTCAGCTACTTTTATTTTCTGTGTTT..... 375
4006 TCTCAGCTACTTNNATTTCTTGTGTTTGGGGAAGATGATCAATCCTTAGT 4055
376TGGTGACATTAgC 388
4106 GTGATTGTGAATCTCCATGATATCTCTTAATATTCTCAGGTGACATTAGC 4155
389 ATTGTATGTTACTGGGCGAAGTtCCTCgAATAAgGtTctTCCCTAAAGC 438
4156 ATTGTATGTTACTGGGCGAAGTTCCTCGAATAAGGGTTCTTCCCTAAAGC 4205
439 CTCATGTCTCagGAAATCTTGTACATGACT..... 468
4206 CTCATGTCTCAGGAAATCTTGTACATGACTGGTACTAACATAATACAATT 4255
469GGTGGTTTGGGAATACAGCTGAATCCTCAGT 498
4556 ATTGTTTGTTCATGTGATAGGTGGTTTGGGAATACAGCTGAATCCTCAGT 4605
499 TTATGAGCATTGATCTCAA..... 517
4606 TTATGAGCATTGATCTCAAGTAATCCATTTTCTGTTTTTCTTCTATTT 4655
518GTTTTTCTTTGTCAGAGCCGGGATGATGGGATGG 551
4756 GTTATTCCTTGTTAGGTTTTCTTTGTCAGAGCCGGGATGATGGGATGG 4805
552 CTGCTTATCAATCTCTCTATTCTGGCAAAAAGTGTGCAGGATGGTTCCTT 601
4806 CTGCTTATCAATCTCTCTATTCTGGCAAAAAGTGTGCAGGATGGTTCCTT 4855
602 GAGTCAGTCGATGATTCTTTACCAGATCTTCTGTSC..... 637
4856 GAGTCAGTCGATGA.TCTTTACCAGATCTTCTGTGCGGTAAATTTGGTTT 4904

FIGURE 16 / SHEET 3

638GTTATATATATTGGACTACTTTGTTTCAT 665
|||||
4955 TTTGATTAGTTTATAAATTGCAGTTATATATATTGGA.TACTTTGTTTCAT 5003
666 GAAGaATACATGACCTCTAC..... 685
|||||
5004 GAAGAATACATGACCTCTACGTAAGTTCATGGCGTGTTAAGGAAACACAT 5053
.
.
.
686GTGGGACATAAATTGCAGAGAGACTAgGCTTCATGCTAGT 724
|||||
5104 ACTCTTTTCAGGTGGGACATAAATTGCAGAGAGACTAGGCTTCATGCTAGT 5153
725 GTTTGGAGATCTCCTGTGGATTCTTTCACTTTTAGCATT..... 764
|||||
5154 GTTTGGAGATCTCCTGTGGATTCTTTCACTTTTAGCATTTCAGGCATGTA 5203
.
.
.
765CAGGGCTGGTGGC 777
|||||
5304 TTGATCACTCTGTGACTGAATTGTCTTCTCTCTGTTACAGGGCTGGTGGC 5353
778 TTTTGCACAACAAAGTAgAACTAACAGTTCCTGCGATTGTAGTCAATTGC 827
|||||
5354 TTTTGCACAACAAAGTAGAACTAACAAATTCTGCGATTGTAGTCAATTGC 5403
828 CTTGTCTTCTTGATAG..... 843
|||||
5404 CTTGTCTTCTTGATAGGGTAAGTTCGTAGACATGGGGTTATTTTCCATTTC 5453
844GGTACATG 851
|||||
5454 TTACATATCTACACTAAGAAACCCACTATTTCTTCTTTGGCAGGTACATG 5503
852 GTTTTTTCGAGgGAgCTAACAAACAAAAACATATCTTTAAGAAGAACCCAAA 901
|||||
5504 GTTTTTTCGAGGAGCTAACAAACAAAAACATATCTTTAAGAAGAACCCAAA 5553
902 AACACCAATATGGGGCAAGCCTCCAGTGGTAGTTGGTGGAAAGTTACTGG 951
|||||
5554 AACACCAATATGGGGCAAGCCTCCAGTGGTAGTTGGTGGAAAGTTACTGG 5603
952 TTTcAGGCTATT..... 963
|||||
5604 TTTcAGGCTATTGGSTATGTTATATTTATCTCTCTCTGTTTCTTTGCTTGG 5653
.
.
.
964GGGGAATTGCAAGGCACtGTAAATTACCTTGGCGACTTGATGCTT 1007
|||||
5754 ATGACAGGGGAATTGCAAGGCACtGTAAATTACCTTGGCGACTTGATGCTT 5803
1008 GCTCTGTCTTCAGTTTGGCCATGTGGAATA..... 1037

FIGURE 16 / SHEET 4

|||||
5804 GCTCTGTCCTTCAGTTTGCCATGTGGAATAAGGTACTCCTNCTGCTTGAG 5853
.
.
.
1038AGTTCTCCGGTTCCATATTTCTACCCGATAT 1068
|||||
6004 TTTCGCTCTNGTGATTCTCAGTTCTCCGGTTCCATATTTCTACCCGATAT 6053
1069 ACCTTCTGATACTATTGATATGGAGAGAACGAAGAGACGAGGTTTCGATGT 1118
|||||
6054 ACCTGCTGATACTATTGATATGGAGAGAACGAAGAGACGAAGTTTCGATGT 6103
1119 GCAGAGAAGTACAAGGAGATATGGGCAGAGTATCTTAGACTTGTCCCCTG 1168
|||||
6104 GCAGAGAAGTACNAGGAGATATGGGCAGAGTATCTTAGACTTGTCCCCTG 6153
1169 GAGAATACTTCCTTATGTTTATTAGATGTGCCAAGAGCCAAGTCATGAAT 1218
|||||
6154 GAGAATACTTCCTTATGTTTATTAGATGTGCCAAGAGCCAATTCATGAAT 6203
1219 CCTTTCAGATTCACCTCTTGTGCTTATTTTTTCCATAA..... 1258
|||||
6204 CCTTTCAGATTCATCCTCTTGTGCTTATTTTTTTCATTAAATGTGACNTG 6253
1259TCTTGTTTTATTTTAGCAATGCTCGAATTGAACTTTGTAG 1299
|| | : | :||| ||||| | |||||
6254 AAATGATCCCATTTATNGCCTNTTATCAATGCTTG.ATTGAACTTTGTAG 6302
1300 TACACTTTTGAAAAATAACTTCAGTCCTT 1328
|||||
6303 TACACGTTTG.AGAATTACTTCAGTCCTT 6330

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/10644

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12N 9/02, 15/53, 15/63, 1/21, 15/09, 15/10; C12Q 1/68, 1/26

US CL : 526/23.2, 23.6; 435/189, 252.3, 320.1, 419, 6, 91.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 526/23.2, 23.6; 435/189, 252.3, 320.1, 419, 6, 91.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, SCISEARCH, LIFESCI, BIOTECHDS, BIOSIS, EMBASE, CAPLUS, NTIS, WPI, CABA, AGRICOLA
search terms: sterol reductase#

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	TATON M. Microsomal $\Delta^{8,14}$ -Sterol Δ^{14} -Reductase in Higher Plants. Eur. J. Biochem. November 1989. Vol. 185. pages 605-614. see entire document.	1-18, 24, 25
Y	GOAD L.J. Application of Sterol Synthesis Inhibitors to Investigate the Sterol Requirements of Protozoa and Plants. Biochem Soc. Trans. 1990. Vol. 18 pages 63-65. see entire document.	1-18, 24, 25
A	LAI, M.H. The Identification of a Gene family in the Saccharomyces cerevisiae Ergosterol Biosynthesis Pathway. Gene 1994. Vol. 140. pages 41-49.	1-18, 24, 25



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

20 AUGUST 1997

Date of mailing of the international search report

04 SEP 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

REBECCA PROBY

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/10644

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-18, 24 and 25

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/10644

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-18, 24 and 25, drawn to C-14 sterolreductase, DNA therefore, and methods of producing C-14 sterol reductase.

Group II, claim 19, drawn to C-14 sterol reductase antibodies.

Group III, claims 20-23, 30 and 31, drawn to transgenic plants.

Group IV, claims 26-29, drawn to a method for reducing C-14 sterol reductase expression.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The inventions of Groups I-III comprise different products with unrelated chemical structures. The method of group IV lacks a corresponding special technical feature because it neither makes nor uses the products of any of Groups I-III.

**CORRECTED
VERSION***

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 9/02, 15/53, 15/63, 1/21, 15/09, 15/10, C12Q 1/68, 1/26	A1	(11) International Publication Number: WO 97/48793 (43) International Publication Date: 24 December 1997 (24.12.97)
(21) International Application Number: PCT/US97/10644 (22) International Filing Date: 20 June 1997 (20.06.97) (30) Priority Data: 60/022,086 21 June 1996 (21.06.96) US (71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, MA 02114 (US). (72) Inventors: JANG, Jyan-Chyun; The General Hospital Corpora- tion, 55 Fruit Street, Boston, MA 02114 (US). SHEEN, Jen; The General Hospital Corporation, 55 Fruit Street, Boston, MA 02114 (US). (74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110-2214 (US).		(81) Designated States: AU, CA, CN, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>
(54) Title: PLANT STEROL REDUCTASES AND USES THEREOF (57) Abstract Disclosed are plant sterol biosynthetic enzymes, genes, and their uses.		

* (Referred to in PCT Gazette No. 11/1998, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakistan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

PLANT STEROL REDUCTASES AND USES THEREOF

5

Background of the Invention

This application relates to plant sterol biosynthetic enzymes, genes, and their uses.

Plant sterols belong to a large group of secondary compounds known as
10 terpenes or isoprenoids. Sterol biosynthesis in plants generally involves a series of different enzymatic steps in the isoprenoid pathway that result in the formation of a variety of sterol end products (Benveniste *Ann. Rev. Biochem.* 37:275, 1986). Although such sterol compounds have been identified in higher plants, their function in plant growth and development is poorly understood.

15 One such plant sterol, brassinolide, that belongs to a class of sterols referred to as brassinosteroids (BR), was first discovered in the pollen of *Brassica napus* (Grove et. al., *Nature* 281: 216, 1979). Brassinosteroids are growth-promoting natural products having structural similarities to animal steroid hormones. The wide distribution of brassinosteroids in the plant kingdom, their effect on cell proliferation
20 and elongation, and their interactions with other plant hormones (e.g., cytokinins), have indicated that these compounds are plant-growth regulators. Brassinosteroids are thought to promote hypocotyl elongation, leaf unrolling, and xylem differentiation. In addition, such compounds are also believed to be involved in de-etiolation of cotyledons, root elongation, radial growth, and anthocyanin formation.

25 The function of plant sterol growth regulators, such as BR, in relationship to other classes of plant growth regulators such as auxin, gibberellin, abscisic acid, and cytokinin, during plant development also needs to be evaluated. For example, the growth regulator, cytokinin, is known to affect a variety of developmental processes including photomorphogenesis, chloroplast biogenesis and maintenance, apical

- 2 -

dominance, and senescence. In addition, this growth regulator is thought to antagonize BR's ability to promote hypocotyl elongation and cotyledon de-etiolation.

Summary of the Invention

5 In general, the invention features a substantially pure plant C-14 sterol reductase polypeptide. Preferably, the C-14 sterol reductase polypeptide includes an amino acid sequence substantially identical to the sequence shown in Fig. 14 (SEQ ID NO: 1); and is from a dicot (for example, a crucifer or a solanaceous plant), monocot, gymnosperm, or an alga.

10 In related aspects, the invention features purified DNA that includes a sequence encoding a C-14 sterol reductase polypeptide (for example, a sequence substantially identical to the DNA sequence shown in Fig. 14; SEQ ID NO: 2; or a DNA sequence that encodes a C-14 sterol reductase polypeptide which has an amino acid sequence substantially identical to that shown in Fig. 14; SEQ ID NO: 1). The invention also
15 features a vector and a cell, each of which includes purified DNA encoding a C-14 sterol reductase polypeptide; and a method of producing a recombinant C-14 sterol reductase polypeptide involving providing a cell (for example, a plant cell) transformed with purified DNA encoding a C-14 sterol reductase polypeptide positioned for expression in the cell, culturing the transformed cell under conditions
20 for expressing the DNA, and isolating the recombinant C-14 sterol reductase polypeptide. The invention further features recombinant C-14 sterol reductase produced by such expression of a purified DNA, and an isolated antibody that specifically recognizes and binds a plant C-14 sterol reductase polypeptide.

25 In addition, the invention features nucleotide sequences that hybridize to a C-14 sterol reductase gene (including the coding sequence of such a gene and its complement) and that encode a C-14 sterol reductase polypeptide. Furthermore, the invention includes oligonucleotide probes that detect a C-14 sterol reductase gene or functional equivalents thereof in a plant (for example, dicots (such as solanaceous and

- 3 -

cruciferous plants), monocots, gymnosperms, and algae). Such probes are useful to isolate DNA sequences that encode C-14 sterol reductases from other plants. In one particular example, oligonucleotides may be designed based on a C-14 sterol reductase sequence disclosed herein and used as hybridization probes or as primers in
5 polymerase chain reactions (PCR). Conserved regions in the C-14 sterol reductase gene are useful in the design of such primers to facilitate the recovery of C-14 sterol reductases from other related and unrelated plants.

In yet other related aspects, the invention features a transgenic plant (or seeds or cells thereof) containing DNA encoding a C-14 sterol reductase polypeptide
10 integrated into the genome of the plant, where the DNA is expressed in the transgenic plant, resulting in the production of a C-14 sterol reductase polypeptide.

In still another aspect, the invention features a method for reducing the level of a plant C-14 sterol reductase polypeptide in a transgenic plant cell. This method generally involves expressing in the transgenic plant cell an antisense C-14 sterol
15 reductase polypeptide nucleic acid sequence. In general, such an antisense C-14 sterol reductase nucleic acid sequence is encoded by a transgene integrated into the genome of the transgenic plant cell and is based on the nucleotide sequence that is shown in Fig. 14 (SEQ ID NO: 2) or Fig. 15. (SEQ ID NO: 3). In preferred embodiments, the plant cell expressing an antisense C-14 sterol reductase nucleic acid sequence is a
20 dicot (for example, crucifer), monocot, gymnosperm, or algal cell. In yet other preferred embodiments, the method involves growing a transgenic plant from the transgenic plant cell, whereby the level of the C-14 sterol reductase polypeptide is reduced in the transgenic plant.

In other related aspects, the invention features a plant cell expressing an
25 antisense C-14 sterol reductase nucleic acid sequence and a plant expression vector that includes an antisense C-14 sterol reductase nucleic acid sequence, where the antisense sequence is operably linked to an expression control region.

- 4 -

In another aspect, the invention features a method for increasing the level of a C-14 sterol reductase in a transgenic plant cell. This method involves expressing in the transgenic plant cell a C-14 sterol reductase polypeptide nucleic acid sequence. Preferably, the method utilizes a C-14 sterol reductase nucleic acid sequence that is substantially identical to the nucleotide sequence that is shown Fig. 14 (SEQ ID NO: 2). In preferred embodiments, the plant cell expressing a C-14 sterol reductase polypeptide nucleic acid sequence is a dicot (for example, a crucifer), monocot, gymnosperm, or algal cell.

In another aspect, the invention features a transgenic plant having a knockout mutation in DNA encoding a plant C-14 sterol reductase polypeptide. Such knockout genes are constructed according to conventional methods (e.g., Lee et al. *Plant Cell* 2: 415, 1990; Miao and Lam, *Plant J.* 7: 359, 1995).

By "plant C-14 sterol reductase" is meant an amino acid sequence that catalyzes the reduction of any sterol precursor having a C14=C15 double bond, for example, as described by Benveniste, *Annu. Rev. Biochem.* 37: 275, 1986. Preferably, such a polypeptide has an amino acid sequence which is at least 30%, preferably 40%, and most preferably 50% or even 80-95% identical to the amino acid sequence of the C-14 sterol reductase polypeptide shown in Fig. 14 (SEQ ID NO: 1). The length of comparison of amino acid sequences will generally be at least 16 amino acids, preferably at least 20 amino acids, more preferably at least 25 amino acids, and most preferably at least 35 amino acids.

By "polypeptide" or "protein" is meant any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation).

By a "substantially identical" polypeptide sequence is meant an amino acid sequence that differs only by conservative amino acid substitutions, for example, substitution of one amino acid for another of the same class (e.g., valine for glycine, arginine for lysine, etc.) or by one or more non-conservative substitutions, deletions,

or insertions, located at positions of the amino acid sequence that do not destroy the function of the polypeptide (assayed, for example, as described herein).

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group
5 (University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705), BLAST, or PILEUP/PRETTYBOX programs). Such software matches sequences by assigning degrees of homology to various substitutions, deletions, and other modifications.

By "substantially pure polypeptide" is meant a polypeptide preparation that is at
10 least 60% by weight (dry weight) the compound of interest, for example, the C-14 sterol reductase polypeptide or C-14 sterol reductase-specific antibody. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight the compound of interest. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel
15 electrophoresis, or HPLC analysis.

By "purified DNA" is meant DNA that is not immediately contiguous with both of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally-occurring genome of the organism from which it is derived. The term therefore includes, for example, a recombinant DNA that is
20 incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote, or that exists as a separate molecule (for example, a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding one or more additional amino
25 acids.

By a "substantially identical" nucleic acid is meant a nucleic acid sequence that encodes a polypeptide differing only by conservative amino acid substitutions, for example, substitution of one amino acid for another of the same class (e.g., valine for

- 6 -

glycine, arginine for lysine, etc.) or by one or more non-conservative substitutions, deletions, or insertions, located at positions of the amino acid sequence that do not destroy the function of the polypeptide (assayed, for example, as described herein). Again, the encoded sequence is at least 30%, more preferably 40%, and most preferably 50%, or even 80 to 95% identical at the amino acid level to the sequence of Fig. 14 (SEQ ID NO: 1). Thus, when nucleic acid sequences are compared, a "substantially identical" nucleic acid sequence is one which is at least 30%, more preferably 40%, and most preferably 50%, or even 80 to 95% identical to the sequence of Fig. 14 (SEQ ID NO: 2). The length of nucleic acid sequence comparison will generally be at least 30 nucleotides, preferably at least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably 110 nucleotides. Again, identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705).

By "isolated antibody" is meant antibody that is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, antibody.

By "specifically binds" is meant an antibody that recognizes and binds a C-14 sterol reductase polypeptide but which does not substantially recognize and bind other molecules in a sample (e.g., a biological sample) which naturally includes a C-14 sterol reductase. An antibody which "specifically binds" a C-14 sterol reductase is sufficient to detect a C-14 sterol reductase product in such a biological sample using one or more of the standard immunological techniques available to those in the art (for example, Western blotting or immunoprecipitation).

By "an antisense C-14 sterol reductase sequence" is meant a nucleotide sequence that is complementary to a plant C-14 sterol reductase messenger RNA. In general, such an antisense sequence will usually be at least 15 nucleotides, preferably

- 7 -

about 15-200 nucleotides, and more preferably 200-2,000 nucleotides in length. The antisense sequence may be complementary to all or a portion of the plant C-14 sterol reductase mRNA nucleotide sequence, and, as appreciated by those skilled in the art, the particular site or sites to which the antisense sequence binds as well as the length of the antisense sequence will vary, depending upon the degree of inhibition desired and the uniqueness of the antisense sequence. By binding to the appropriate target sequence, an RNA-RNA, DNA-DNA, or RNA-DNA duplex is formed. A transcriptional construct expressing a plant C-14 sterol reductase antisense nucleotide sequence includes, in the direction of transcription, a promoter, the sequence coding for the antisense RNA on the sense strand, and a transcriptional termination region. Antisense C-14 sterol reductase sequences may be constructed and expressed as described herein or as described, for example, in van der Krol et al., *Gene* 72: 45, 1988; Rodermel et al., *Cell* 55: 673, 1988; Mol et al., *FEBS Lett.* 268: 427, 1990; Weigel and Nilsson, *Nature* 377: 495, 1995; Cheung et al., *Cell* 82, 383, 1995; and U.S. Pat. No. 5,107,065. In addition, C-14 sterol reductase antisense sequences are useful for the formation of triple helices, where the antisense sequence is bound to a DNA duplex. By binding to the target nucleic acid, C-14 sterol reductase antisense sequences can inhibit the function of the target nucleic acid. This results, for example, in the blocking of transcription, processing of poly A+ addition, replication, translation, or promoting inhibitory mechanisms of the cell, such as RNA degradation. The triple helix-forming and antisense C-14 sterol reductase sequences are useful for selectively suppressing certain cellular functions that are associated with C-14 sterol reductase activity.

By a "transformed cell" is meant a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule encoding (as used herein) a C-14 sterol reductase polypeptide (for example, a substantially identical DNA encoding the C-14 sterol reductase shown in Fig. 14 (SEQ ID NO: 2)).

- 8 -

By "positioned for expression" is meant that the DNA molecule is positioned adjacent to a DNA sequence which directs transcription and translation of the sequence (for example, facilitates the production of, for example, a plant C-14 sterol reductase polypeptide such as the amino acid sequence shown in Fig. 14 (SEQ ID NO: 1)), or an RNA molecule (for example, an antisense RNA).

By "promoter" is meant a minimal sequence sufficient to direct transcription. Included in the invention are promoter elements that are sufficient to render promoter-dependent gene expression controllable for cell-, tissue-, or organ-specific gene expression, or elements that are inducible by external signals or agents (for example, light-, pathogen-, wound-, stress-, or hormone-inducible elements); such elements may be located in the 5' or 3' regions of the native gene or engineered into a transgene construct.

By "operably linked" is meant that a gene and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (for example, transcriptional activator proteins) are bound to the regulatory sequence(s).

By "crucifer" is meant any plant that is classified within the Cruciferae family as commonly described in, e.g., Gray's Manual of Botany American Book Company, N.Y., 1950; *Hortus Third: A Concise Dictionary of Plants Cultivated in the U.S. and Canada*, Macmillan, 1976; or Simmons, N.W., *Evolution of Crop Plants*, 1986. The Cruciferae include many agricultural crops, including, but not limited to, broccoli, cabbage, brussel sprouts, rapeseed, kale, Chinese kale, cauliflower, horseradish, and *Arabidopsis*.

By "plant cell" is meant any self-propagating cell bounded by a semi-permeable membrane and containing a plastid. Such a cell also requires a cell wall if further propagation is desired. Plant cell, as used herein includes, without limitation, algae, cyanobacteria, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores.

By "transgene" is meant any piece of DNA that is inserted by artifice into a cell, and becomes part of the genome of the organism which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (i.e., foreign) to the transgenic organism, or may represent a gene homologous to an
5 endogenous gene of the organism.

By "transgenic" is meant any cell that includes a DNA sequence which is inserted by artifice into a cell and becomes part of the genome of the organism which develops from that cell. As used herein, the transgenic organisms are generally transgenic plants and the DNA (transgene) is inserted by artifice into the nuclear or
10 plastidic genomes.

Other features and advantages of the invention will be apparent from the following detailed description thereof, and from the claims.

Detailed Description

The drawings will first be described.

Drawings

Figs. 1A-1B are photographs showing that the *ell* mutant morphology was phenocopied by treating wild-type seedlings with 30 μ M of dimethylallylamino-purine (2ip), a synthetic cytokinin, in the dark (**Fig. 1A**) or light (**Fig. 1B**). From left to right in both **Figs. 1A-1B**: wild-type plant; wild-type plant + 2ip; *ell*; and *ell* + 2ip.

Fig. 2 is a photograph illustrating the constitutive photomorphogenesis of *ell* seedling development in the dark. Wild-type (left) and *ell* (right) seedlings were grown in the dark for twenty-one days on Murashige-Skoog (MS) plates containing two percent sucrose.

Figs. 3A-3B are photographs showing that the rosette leaves of the *ell* plant (**Fig. 3B**) are darker green in color than those of the wild-type plant (**Fig. 3A**).

Fig. 4 is a photograph illustrating that an *ell* mutant has reduced apical dominance in comparison to a wild-type plant. Six-week-old wild-type (left) and ten-week-old *ell* (right) plants were grown in the greenhouse.

Fig. 5 is a photograph showing that *ell* mutants (right) exhibit irregular, thickened cotyledons and hypocotyls, and reduced cotyledon petioles compared to wild-type plants (left).

Figs. 6A-6B are photographs showing abnormal flower development in the *ell* mutant. **Fig. 6A** shows, from left to right, that the sepal, petal, stamen, and carpel are shorter in *ell* (lower row) than wild-type (upper row) plants. **Fig. 6B** shows, from left to right, the top and side view of wild-type (left) and *ell* (right) flowers.

Figs. 7A-7F are photographs showing embryo development in *ell* and wild-type plants. Wild-type and *ell* plants are shown in the left and right of each photograph, respectively. **Fig. 7A** shows *ell* embryo development at the 32- to 64-cell stage, and **Fig. 7B** shows that, when wild-type embryos have reached the heart stage, *ell* embryos are only at the globular stage. As shown in **Fig. 7C** and **Fig. 7D**, when the wild-type embryo reached the torpedo stage, the *ell* mutant embryo was at the heart stage. **Fig. 7E** shows that apical hooks were not formed in *ell* embryos. And

Fig. 7F shows that *ell* seeds desiccated without completing the late stages of embryogenesis.

Figs. 8A-8B are photographs of dry seeds from wild-type (**Fig. 8A**) and *ell* plants (**Fig. 8B**). Reduced seed size, wrinkled seed coat, and precocious germination were observed in *ell* seeds.

Figs. 9A-9F are photographs showing the supernumerary cotyledons that were observed in the *ell* mutant, including one (**Fig. 9A**), two (**Fig. 9B**), three (**Fig. 9C**), four (**Fig. 9D**), five (**Fig. 9E**), and more than six cotyledons (**Fig. 9F**).

Figs. 10A-10B are illustrations showing various aspects of the molecular characterization of the *Arabidopsis thaliana* C-14 sterol reductase gene. **Fig. 10A** is a schematic illustration showing the position of a T-DNA insertion into chromosome 3 of *Arabidopsis*, approximately forty base pairs upstream of the *ELL* gene, and the exon-intron structure of the C-14 sterol reductase gene.

Fig. 10B is a schematic illustration showing the map position of *ELL* on chromosome 3.

Fig. 11 is a schematic illustration showing a comparison of the predicated *ELL* amino acid sequence (designated Ath; SEQ ID NO: 1) with C-14 sterol reductase of *Saccharomyces cerevisiae* (Erg24) and *Schizosaccharomyces pombe* (Pombe), and C-24 sterol reductase of *Sz. pombe* (Sts1) and *S. cerevisiae* (Ygl022).

Fig. 12 is a schematic illustration showing that the predicted *ELL* amino acid sequence (designated Ath; SEQ ID NO: 1) shares homology to human and chicken lamin B receptor.

Figs. 13A-13B are photographs showing that the *ell* phenotype was not corrected by exogenous feeding of brassinolide (1 μ M) in either dark (**Fig. 13A**) or light (**Fig. 13B**). From left to right in **Figs. 13A-13B**: wild-type; *ell*; wild-type + brassinolide; and *ell* + brassinolide.

Fig. 14 is a schematic illustration showing the nucleotide sequence of an *Arabidopsis* C-14 sterol reductase (SEQ ID NO: 2) and its deduced amino acid sequence polypeptide (SEQ ID NO: 1).

Fig. 15 is a schematic illustration showing the genomic nucleotide sequence of an *Arabidopsis* C-14 sterol reductase polypeptide (SEQ ID NO: 3).

Fig. 16 is a schematic illustration showing the sequence comparison between the genomic nucleotide sequence (SEQ ID NO: 3) and cDNA sequences (SEQ ID NO: 2) of an *Arabidopsis* C-14 sterol reductase.

There now follows a description of an *Arabidopsis* mutant, *ell* (extra long life), that displays a life span that is at least three times greater than wild-type plants. The *ell* mutant was isolated by T-DNA tagging methods and was shown to encode a novel C-14 sterol reductase. This example is provided for the purpose of illustrating the invention, and should not be construed as limiting.

Identification and Developmental Effects of the *ell* Mutation By screening for mutants displaying BR deficiency or constitutive cytokinin activity, a recessive mutation causing pleiotropic developmental effects was identified according to conventional methods in an *Arabidopsis* T-DNA insertional mutant collection (Feldmann, *Plant J.* 1:71, 1991; Errampalli et al., *Plant Cell* 3: 149, 1991). This mutant, termed "ell", was found to have a number of developmental abnormalities.

For example, unlike wild-type plants, *ell* mutants displayed constitutive light-morphogenesis (Fig. 2), similar to the *Arabidopsis det2* (Chory et al., *Plant Cell* 3: 445, 1991) and *cpd* (Szekeres et al., *Cell* 85: 171, 1996) mutants. In addition, compared to wild-type plants, *ell* plants had darker green rosette leaves (Figs. 3A-3B), reduced apical dominance (Fig. 4), stunted hairy roots, and irregular hypocotyl and cotyledons (Fig. 5). Furthermore, as shown in Figs. 6A-6B, the *ell* mutant showed reduced and ruffled sepals and petals. The *ell* mutant also showed delayed and altered

embryo development (Figs. 7A-7F) and was found to have reduced fertility, producing wrinkled seeds that precociously germinated (Figs. 8A-8B). In addition, the various phenotypes of *ell* overlapped with *amp-1 (pt1)* (Chaudhury et. al., *Plant J.* 4: 907, 1993) and *häuptling* (Jürgens et al., *Ann. Rev. Genet.* 28: 351, 1994), including
5 supernumerary cotyledons (Figs. 9A-9F).

Finally, as shown in Figs. 1A-1B, the morphology of the T-DNA tagged *ell* mutant seedlings was phenocopied by treating wild-type seedlings with 30 μ M dimethylallylamine purine (2ip), a synthetic cytokinin.

Despite having a number of developmental abnormalities, *ell* mutants were
10 found to have a life span that was at least three times greater than wild-type plants.

Genetic Analysis and Molecular Cloning of *ELL*

Standard segregation analysis indicated that *ell* is a recessive mutation. The T2 population of the transgenic line carrying the *ell* mutant showed a 3:1 Mendelian segregation of the T-DNA using kanamycin resistance (*kan'*) as a selectable marker.

15 Of the *kan'* plants, thirty-three percent showed the *ell* phenotype, indicating that the *ell* mutation was recessive. A T3 population was then generated from selfed T2 *kan'* plants having the wild-type phenotype, and the *kan'* marker showed a 3:1 segregation.

Of the seventy-five percent displaying *kan'*, twenty-five percent showed the *ell* phenotype. Because *ell* homozygous plants were found to be either lethal or sterile,

20 T2 heterozygous *ell* plants were subsequently backcrossed to wild-type plants for additional segregation analysis. The resulting F1 population from this backcross showed a 1:1 segregation of the *kan'* marker; no plants were observed having the *ell* phenotype. The F1 *kan'* individuals of the backcross were then selfed to produce an F2 population. Seventy-five percent of this F2 population was found to be *kan'*, and
25 thirty-three percent of the *kan'* resistant plants showed the *ell* phenotype, confirming the recessive nature of this mutant. Consistent segregation of the *ell* phenotype and *kan'* marker was also observed in a subsequent backcross, further indicating that *ell* was tagged by the T-DNA.

- 14 -

Genomic DNA blot analysis, using an NPTII probe derived from the T-DNA vector, showed that a unique single copy of T-DNA was integrated into the *ell* genome. This result, together with the segregation data described above, further indicated that the *ell* phenotype was associated with the *kan^r* marker, and that the *ell* mutation resulted from a single T-DNA insertion in the *Arabidopsis* genome.

The T-DNA-tagged locus was then isolated by constructing a genomic DNA library from the *ell* mutant and was mapped by hybridization using the NPTII probe. Fig. 10A shows the physical map of the T-DNA tagged locus that was determined by DNA hybridization. One of three genomic clones that were found to hybridize to the NPTII probe was partially sequenced and found to have a complete T-DNA insertion and flanking plant sequences. A segment of this genomic clone containing both T-DNA and plant sequences was then used to screen a genomic library that was prepared from wild-type plants. Two positive clones that were identified in this screen were then sequenced. The genomic nucleotide sequence is presented in Fig. 15 (SEQ ID NO: 3).

The T-DNA-plant DNA insert junctions were also used as probes to screen a cDNA library that was prepared from wild-type plants. One isolated cDNA clone, designated D13, was found to have a nucleotide sequence (SEQ ID NO: 1) that matched the genomic sequences flanking the right T-DNA border. Comparison of the cDNA (Fig. 14) with the genomic DNA sequence (Fig. 15) also revealed that the T-DNA was inserted at a location forty base pairs upstream of the 5' end of the *ELL* cDNA transcript (Fig. 16). The complete genomic fragment covering the cDNA sequence was composed of 14 exons and 13 introns (Fig. 10A). Probes that were prepared from both the cDNA or genomic clone were then used for DNA blot analysis. Results from this analysis confirmed that the *ELL* gene was of plant origin.

We also determined the chromosomal position of *ELL* by standard segregation analysis of restriction fragment length polymorphisms (RFLPs) in recombinant inbred lines (Nam et al., *Plant Cell* 1: 699, 1989; Lister and Dean, *Plant J.* 4: 745, 1993;

Hauge et al., *Plant J.* 3: 745, 1993; Schmidt et al., *Science* 270: 480, 1995; Zachgo et al., *Genomic Res.* 6: 19, 1996). By this analysis, we found that *ELL* is located on chromosome 3 and is flanked by the chromosomal markers by mi456 and g2778 (Fig. 10B).

5

ELL Encodes a Novel C-14 Sterol Reductase

A comparison of the deduced polypeptide sequence of the full-length *ELL* cDNA clone to the GenBank database showed that *ELL* had 35% identity to C-14 sterol reductase (Erg24) in yeast (Lorenz and Parks, *DNA Cell Biol.* 9: 685, 1992; Lai et al., *Gene* 140: 41, 1994) (Fig. 11) and 40% identity to the lamin B receptor (LBR) in humans (Ye and Worman, *J. Biol. Chem.* 269: 11306, 1994) (Fig. 12). In addition, the amino acid sequence of *ELL* predicted several hydrophobic regions and between eight to nine transmembrane domains, consistent with the yeast Erg24 and human LBR. However, *ELL* was observed to lack a basic nucleoplasmic amino-terminal domain of about 200 amino acids that has been identified in human LBR. Database searches also revealed that at least two *Arabidopsis* expression sequence tagged (EST) clones (GenBank accession numbers T45011 and T42407) shared homology to *ELL*. DNA sequencing revealed that T45011 encodes an unknown gene with 60% nucleotide sequence identity to *ELL*. The predicted amino acid sequence of T45011 was also observed to have greater than 50% identity to the yeast ERG24 and human LBR. These results further confirmed that *ELL* is encoded by a gene that is a member of the C-14 sterol reductase gene family. T42407 was found to encode an *Arabidopsis* sterol Δ^7 -reductase (Lecain et al., *J. Biol. Chem.* 271: 10866, 1996) that shares 32% amino acid identity to *ELL*.

10

15

20

25

RNA blot analysis indicated multiple transcripts hybridizing to the full-length *ELL* cDNA.

To determine whether the *ell* mutant phenotype is corrected by exogenous feeding of brassinolide, we germinated *ell* seedlings on agar plates containing 1 μ M

brassinolide or 1 μ M 24-epibrassinolide (Li et al., *Science* 272: 398, 1996). The results of these experiments showed that the presence of brassinolide or 24-epibrassinolide, in the growth medium of *ell* plants did not alter the mutant phenotype (Fig. 13A-13B). Thus, it appears that steroid compounds other than BRs are needed to restore an *ell* mutant to a normal growth and development phenotype, as reflected by the pleiotropic phenotypes such as stunted roots (Fig. 2) and impaired embryogenesis (Fig. 7A-7F).

To confirm that ELL activity was upstream of DET2 in the sterol biosynthesis pathway, a double mutant between *ell* and *det2* was constructed and analyzed. The phenotype of *det2/ell* was indistinguishable from *ell*, further supporting the hypothesis that DET2 was epistatic to ELL.

Isolation of Other C-14 Sterol Reductase cDNAs and Genomic DNAs

Based on the C-14 sterol reductase genes and polypeptides described herein, the isolation of additional plant C-14 sterol reductase coding sequences is made possible using standard strategies and techniques that are well known in the art. For example, using all or a portion of the amino acid sequence of a C-14 sterol reductase polypeptide, one may readily design C-14 sterol reductase-specific oligonucleotide probes, including C-14 sterol reductase degenerate oligonucleotide probes (i.e., a mixture of all possible coding sequences for a given amino acid sequence). These oligonucleotides may be based upon the sequence of either DNA strand and any appropriate portion of the C-14 sterol reductase sequence (for example, Fig. 14; SEQ ID NOS: 2 and 1, respectively; and Fig. 15 (SEQ ID NO: 3). General methods for designing and preparing such probes are provided, for example, in Ausubel et al., 1996, *Current Protocols in Molecular Biology*, Wiley Interscience, New York, and Berger and Kimmel, *Guide to Molecular Cloning Techniques*, 1987, Academic Press, New York. These oligonucleotides are useful for C-14 sterol reductase gene isolation, either through their use as probes capable of hybridizing to C-14 sterol reductase

- 17 -

complementary sequences or as primers for various amplification techniques, for example, polymerase chain reaction (PCR) cloning strategies.

Hybridization techniques and screening procedures are well known to those skilled in the art and are described, for example, in Ausubel et al. (*supra*); Berger and Kimmel (*supra*); and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York. If desired, a combination of different oligonucleotide probes may be used for the screening of a recombinant DNA library. The oligonucleotides may be detectably-labeled using methods known in the art and used to probe filter replicas from a recombinant DNA library. Recombinant DNA libraries are prepared according to methods well known in the art, for example, as described in Ausubel et al. (*supra*), or they may be obtained from commercial sources.

For detection or isolation of closely related C-14 sterol reductase sequences having greater than 80% identity, high stringency conditions are preferably used; such conditions include hybridization at about 65°C and about 50% formamide, a first wash at about 65°C, about 2X SSC, and 1% SDS, followed by a second wash at about 65°C and about 0.1% SDS, and 0.1X SSC. Lower stringency conditions for detecting C-14 sterol reductase genes having about 40-50% sequence identity to the C-14 sterol reductase genes described herein include, for example, hybridization at about 37°C in the absence of formamide, a first wash at about 37°C, about 6X SSC, and about 1% SDS, and a second wash at about 37°C, about 6X SSC, and about 1% SDS. These stringency conditions are exemplary; other appropriate conditions may be determined by those skilled in the art.

As discussed above, C-14 sterol reductase oligonucleotides may also be used as primers in amplification cloning strategies, for example, using PCR. PCR methods are well known in the art and are described, for example, in *PCR Technology*, Erlich, ed., Stockton Press, London, 1989; *PCR Protocols: A Guide to Methods and Applications*, Innis et al., eds., Academic Press, Inc., New York, 1990; and Ausubel et

al. (*supra*). Primers are optionally designed to allow cloning of the amplified product into a suitable vector, for example, by including appropriate restriction sites at the 5' and 3' ends of the amplified fragment (as described herein). If desired, C-14 sterol reductase sequences may be isolated using the PCR "RACE" technique, or Rapid
5 Amplification of cDNA Ends (see, e.g., Innis et al. (*supra*)). By this method, oligonucleotide primers based on a C-14 sterol reductase sequence are oriented in the 3' and 5' directions and are used to generate overlapping PCR fragments. These overlapping 3'- and 5'-end RACE products are combined to produce an intact full-length cDNA. This method is described in Innis et al. (*supra*); and Frohman et al.,
10 *Proc. Natl. Acad. Sci. USA* 85: 8998, 1988.

Alternatively, any plant cDNA expression library may be screened by functional complementation of a yeast C-14 reductase mutant (for example, the *erg24* mutant described by Lorenz and Parks, *DNA Cell Biol.* 9: 685, 1992) according to standard methods.

15 Useful C-14 sterol reductase sequences may be isolated from any appropriate organism. Confirmation of a sequence's relatedness to the C-14 sterol reductase polypeptide family may be accomplished by a variety of conventional methods including, but not limited to, functional complementation assays and sequence comparison. In addition, the activity of any C-14 sterol reductase sequence may be
20 evaluated according to any of the techniques described herein.

C-14 Sterol Reductase Polypeptide Expression

C-14 sterol reductase polypeptides may be produced by transformation of a suitable host cell with all or part of a C-14 sterol reductase cDNA (for example, the
25 cDNA described above) in a suitable expression vehicle or with a plasmid construct engineered for increasing the expression of a C-14 sterol reductase polypeptide (*supra*) *in vivo*.

Those skilled in the field of molecular biology will understand that any of a wide variety of expression systems may be used to provide the recombinant protein. The precise host cell used is not critical to the invention. The C-14 sterol reductase protein may be produced in a prokaryotic host, for example, *E. coli*, or in a eukaryotic host, for example, *Saccharomyces cerevisiae*, mammalian cells (for example, COS 1 or NIH 3T3 cells), or any of a number of plant cells including, without limitation, algae, tree species, ornamental species, temperate fruit species, tropical fruit species, vegetable species, legume species, monocots, dicots, or in any plant of commercial or agricultural significance. Particular examples of suitable plant hosts include, but are not limited to, Conifers, Petunia, Tomato, Potato, Tobacco, *Arabidopsis*, Lettuce, Sunflower, Oilseed rape, Flax, Cotton, Sugarbeet, Celery, Soybean, Alfalfa, *Medicago*, Lotus, *Vigna*, Cucumber, Carrot, Eggplant, Cauliflower, Horseradish, Morning Glory, Poplar, Walnut, Apple, Asparagus, Rice, Maize, Millet, Onion, Barley, Orchard grass, Oat, Rye, and Wheat.

Such cells are available from a wide range of sources including the American Type Culture Collection (Rockland, MD); or from any of a number seed companies, for example, W. Atlee Burpee Seed Co. (Warminster, PA), Park Seed Co. (Greenwood, SC), Johnny Seed Co. (Albion, ME), or Northrup King Seeds (Harstville, SC). Descriptions and sources of useful host cells are also found in Vasil I.K., *Cell Culture and Somatic Cell Genetics of Plants*, Vol I, II, III Laboratory Procedures and Their Applications Academic Press, New York, 1984; Dixon, R.A., *Plant Cell Culture-A Practical Approach*, IRL Press, Oxford University, 1985; Green et al., *Plant Tissue and Cell Culture*, Academic Press, New York, 1987; and Gasser and Fraley, *Science* 244: 1293, 1989.

For prokaryotic expression, DNA encoding a C-14 sterol reductase polypeptide is carried on a vector operably linked to control signals capable of effecting expression in the prokaryotic host. If desired, the coding sequence may contain, at its 5' end, a sequence encoding any of the known signal sequences capable of effecting secretion

of the expressed protein into the periplasmic space of the host cell, thereby facilitating recovery of the protein and subsequent purification. Prokaryotes most frequently used are various strains of *E. coli*; however, other microbial strains may also be used.

Plasmid vectors are used which contain replication origins, selectable markers, and control sequences derived from a species compatible with the microbial host.

Examples of such vectors are found in Pouwels et al. (*supra*) or Ausubel et al. (*supra*).

Commonly used prokaryotic control sequences (also referred to as "regulatory elements") are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences. Promoters commonly used to direct protein expression include the beta-lactamase (penicillinase), the lactose (*lac*) (Chang et al., *Nature* 198: 1056, 1977), the tryptophan (*Trp*) (Goeddel et al., *Nucl. Acids Res.* 8: 4057, 1980), and the *tac* promoter systems, as well as the lambda-derived *P_L* promoter and N-gene ribosome binding site (Simatake et al., *Nature* 292: 128, 1981).

One particular bacterial expression system for C-14 sterol reductase polypeptide production is the *E. coli* pET expression system (Novagen, Inc., Madison, WI). According to this expression system, DNA encoding a C-14 sterol reductase polypeptide is inserted into a pET vector in an orientation designed to allow expression. Since the C-14 sterol reductase gene is under the control of the T7 regulatory signals, expression of C-14 sterol reductase is induced by inducing the expression of T7 RNA polymerase in the host cell. This is typically achieved using host strains which express T7 RNA polymerase in response to IPTG induction. Once produced, recombinant C-14 sterol reductase polypeptide is then isolated according to standard methods known in the art, for example, those described herein.

Another bacterial expression system for C-14 sterol reductase polypeptide production is the pGEX expression system (Pharmacia). This system employs a GST gene fusion system which is designed for high-level expression of genes or gene fragments as fusion proteins with rapid purification and recovery of functional gene

products. The protein of interest is fused to the carboxyl terminus of the glutathione S-transferase protein from *Schistosoma japonicum* and is readily purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4B. Fusion proteins can be recovered under mild conditions by elution with glutathione.

5 Cleavage of the glutathione S-transferase domain from the fusion protein is facilitated by the presence of recognition sites for site-specific proteases upstream of this domain. For example, proteins expressed in pGEX-2T plasmids may be cleaved with thrombin; those expressed in pGEX-3X may be cleaved with factor Xa.

For eukaryotic expression, the method of transformation or transfection and the
10 choice of vehicle for expression of the C-14 sterol reductase polypeptide will depend on the host system selected. Transformation and transfection methods are described, e.g., in Ausubel et al. (*supra*); Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989; Gelvin et al., *Plant Molecular Biology Manual*, Kluwer Academic Publishers, 1990; Kindle, K., *Proc. Natl. Acad. Sci., U.S.A*
15 87: 1228, 1990; Potrykus, I., *Annu. Rev. Plant Physiol. Plant Mol. Biology* 42: 205, 1991; and BioRad (Hercules, CA) Technical Bulletin #1687 (Biolistic Particle Delivery Systems). Expression vehicles may be chosen from those provided, e.g., in *Cloning Vectors: A Laboratory Manual* (P.H. Pouwels et al., 1985, Supp. 1987); Gasser and Fraley (*supra*); Clontech Molecular Biology Catalog (Catalog 1992/93
20 Tools for the Molecular Biologist, Palo Alto, CA); and the references cited above.

Most preferably, an C-14 sterol reductase polypeptide is produced by a stably-transfected plant cell line, a transiently-transfected plant cell line, or by a transgenic plant. A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants are available to the public; such vectors are
25 described in Pouwels et al. (*supra*), Weissbach and Weissbach (*supra*), and Gelvin et al. (*supra*). Methods for constructing such cell lines are described in, e.g., Weissbach and Weissbach (*supra*), and Gelvin et al. (*supra*). Typically, plant expression vectors include (1) a cloned plant gene under the transcriptional control of 5' and 3' regulatory

- 22 -

sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (for example, one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Alternatively, the C-14 sterol reductase polypeptide may be produced using a transient expression system (e.g., the maize transient expression system described by Sheen, *Plant Cell* 2: 1027, 1990).

Once the desired C-14 sterol reductase nucleic acid sequences is obtained, it may be manipulated in a variety of ways known in the art. For example, where the sequence involves non-coding flanking regions, the flanking regions may be subjected to mutagenesis.

The C-14 sterol reductase DNA sequence of the invention may, if desired, be combined with other DNA sequences in a variety of ways. The C-14 sterol reductase DNA sequence of the invention may be employed with all or part of the gene sequences normally associated with the C-14 sterol reductase protein. In its component parts, a DNA sequence encoding a C-14 sterol reductase protein is combined in a DNA construct having a transcription initiation control region capable of promoting transcription and translation in a host cell.

In general, the constructs will involve regulatory regions functional in plants which provide for modified production of C-14 sterol reductase protein as discussed herein. The open reading frame coding for the C-14 sterol reductase protein or functional fragment thereof will be joined at its 5' end to a transcription initiation regulatory region such as the sequence naturally found in the 5' upstream region of the C-14 sterol reductase structural gene. Numerous other transcription initiation regions are available which provide for constitutive or inducible regulation.

- 23 -

For applications where developmental, cell, tissue, hormonal, or environmental expression is desired, appropriate 5' upstream non-coding regions are obtained from other genes, for example, from genes regulated during meristem development, seed development, embryo development, or leaf development.

5 Regulatory transcript termination regions may also be provided in DNA constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the C-14 sterol reductase protein or any convenient transcription termination region derived from a different gene source. The transcript termination region will contain preferably at least 1-3 kb of sequence 3' to the
10 structural gene from which the termination region is derived. Plant expression constructs having C-14 sterol reductase as the DNA sequence of interest for expression (in either the sense or antisense orientation) may be employed with a wide variety of plant life, particularly plant life involved in the production of storage reserves (for example, those involving carbon and nitrogen metabolism). Such
15 genetically-engineered plants are useful for a variety of industrial and agricultural applications as discussed below. Importantly, this invention is applicable to dicotyledons and monocotyledons, and will be readily applicable to any new or improved transformation or regeneration method.

An example of a useful plant promoter according to the invention is a
20 caulimovirus promoter, for example, a cauliflower mosaic virus (CaMV) promoter. These promoters confer high levels of expression in most plant tissues, and the activity of these promoters is not dependent on virally encoded proteins. CaMV is a source for both the 35S and 19S promoters. In most tissues of transgenic plants, the CaMV 35S promoter is a strong promoter (see, e.g., Odell et al., *Nature* 313: 810 1985). The
25 CaMV promoter is also highly active in monocots (see, e.g., Dekeyser et al., *Plant Cell* 2: 591, 1990; Terada and Shimamoto, *Mol. Gen. Genet.* 220: 389, 1990). Moreover, activity of this promoter can be further increased (i.e., between 2-10 fold) by duplication of the CaMV 35S promoter (see e.g., Kay et al., *Science* 236: 1299,

1987; Ow et al., *Proc. Natl. Acad. Sci., U.S.A.* 84: 4870, 1987; and Fang et al., *Plant Cell* 1: 141, 1989).

Other useful plant promoters include, without limitation, the nopaline synthase promoter (An et al., *Plant Physiol.* 88: 547, 1988) and the octopine synthase promoter (Fromm et al., *Plant Cell* 1: 977, 1989).

For certain applications, it may be desirable to produce the C-14 sterol reductase gene product in an appropriate tissue, at an appropriate level, or at an appropriate developmental time. For this purpose, there are an assortment of gene promoters, each with its own distinct characteristics embodied in its regulatory sequences, shown to be regulated in response to the environment, hormones, and/or developmental cues. These include gene promoters that are responsible for heat-regulated gene expression (see, e.g., Callis et al., *Plant Physiol.* 88: 965, 1988; Takahashi and Komeda, *Mol. Gen. Genet.* 219: 365, 1989; and Takahashi et al. *Plant J.* 2: 751, 1992), light-regulated gene expression (e.g., the pea *rbcS-3A* described by Kuhlemeier et al., *Plant Cell* 1: 471, 1989; the maize *rbcS* promoter described by Schäffner and Sheen, *Plant Cell* 3: 997, 1991; or the chlorophyll a/b-binding protein gene found in pea described by Simpson et al., *EMBO J.* 4: 2723, 1985), hormone-regulated gene expression (for example, the abscisic acid (ABA) responsive sequences from the *Em* gene of wheat described by Marcotte et al., *Plant Cell* 1: 969, 1989; the ABA-inducible HVA1 and HVA22, and rd29A promoters described for barley and *Arabidopsis* by Straub et al., *Plant Cell* 6: 617, 1994, Shen et al., *Plant Cell* 7: 295, 1995; and wound-induced gene expression (for example, of *wun1* described by Siebertz et al., *Plant Cell* 1: 961, 1989), or organ-specific gene expression (for example, of the tuber-specific storage protein gene described by Roshal et al., *EMBO J.* 6: 1155, 1987; the 23-kDa zein gene from maize described by Schernthaner et al., *EMBO J.* 7: 1249, 1988; or the French bean β -phaseolin gene described by Bustos et al., *Plant Cell* 1: 839, 1989).

- 25 -

Plant expression vectors may also optionally include RNA processing signals, e.g, introns, which have been shown to be important for efficient RNA synthesis and accumulation (Callis et al., *Genes and Dev.* 1: 1183, 1987). The location of the RNA splice sequences can dramatically influence the level of transgene expression in plants.

5 In view of this fact, an intron may be positioned upstream or downstream of a C-14 sterol reductase polypeptide-encoding sequence in the transgene to modulate levels of gene expression.

In addition to the aforementioned 5' regulatory control sequences, the expression vectors may also include regulatory control regions which are generally
10 present in the 3' regions of plant genes (Thornburg et al., *Proc. Natl. Acad. Sci. U.S.A.* 84: 744, 1987; An et al., *Plant Cell* 1: 115, 1989). For example, the 3' terminator region may be included in the expression vector to increase stability of the mRNA. One such terminator region may be derived from the PI-II terminator region of potato. In addition, other commonly used terminators are derived from the octopine or
15 nopaline synthase signals.

The plant expression vector also typically contains a dominant selectable marker gene used to identify those cells that have become transformed. Useful selectable genes for plant systems include genes encoding antibiotic resistance genes, for example, those encoding resistance to hygromycin, kanamycin, bleomycin, G418,
20 streptomycin, or spectinomycin. Genes required for photosynthesis may also be used as selectable markers in photosynthetic-deficient strains. Alternatively, the green-fluorescent protein from the jellyfish *Aequorea victoria* may be used as a selectable marker (Sheen et al., *Plant J.* 8:777, 1995; Chiu et al., *Current Biology* 6: 325, 1996). Finally, genes encoding herbicide resistance may be used as selectable markers; useful
25 herbicide resistance genes include the *bar* gene encoding the enzyme phosphinothricin acetyltransferase and conferring resistance to the broad spectrum herbicide Basta® (Hoechst AG, Frankfurt, Germany).

- 26 -

Efficient use of selectable markers is facilitated by a determination of the susceptibility of a plant cell to a particular selectable agent and a determination of the concentration of this agent which effectively kills most, if not all, of the transformed cells. Some useful concentrations of antibiotics for tobacco transformation include, e.g., 75-100 µg/ml (kanamycin), 20-50 µg/ml (hygromycin), or 5-10 µg/ml (bleomycin). A useful strategy for selection of transformants for herbicide resistance is described, e.g., by Vasil et al., *supra*.

It should be readily apparent to one skilled in the art of molecular biology, especially in the field of plant molecular biology, that the level of gene expression is dependent, not only on the combination of promoters, RNA processing signals, and terminator elements, but also on how these elements are used to increase the levels of selectable marker gene expression.

Plant Transformation

Upon construction of the plant expression vector, several standard methods are available for introduction of the vector into a plant host, thereby generating a transgenic plant. These methods include (1) *Agrobacterium*-mediated transformation (*A. tumefaciens* or *A. rhizogenes*) (see, e.g., Lichtenstein and Fuller In: *Genetic Engineering*, vol 6, PWJ Rigby, ed, London, Academic Press, 1987; and Lichtenstein, C.P., and Draper, J., In: *DNA Cloning*, Vol II, D.M. Glover, ed, Oxford, IRI Press, 1985)), (2) the particle delivery system (see, e.g., Gordon-Kamm et al., *Plant Cell* 2: 603 (1990); or BioRad Technical Bulletin 1687, *supra*), (3) microinjection protocols (see, e.g., Green et al., *supra*), (4) polyethylene glycol (PEG) procedures (see, e.g., Draper et al., *Plant Cell Physiol.* 23: 451, 1982; or e.g., Zhang and Wu, *Theor. Appl. Genet.* 76: 835, 1988), (5) liposome-mediated DNA uptake (see, e.g., Freeman et al., *Plant Cell Physiol.* 25: 1353, 1984), (6) electroporation protocols (see, e.g., Gelvin et al., *supra*; Dekeyser et al., *supra*; Fromm et al., *Nature* 319: 791, 1986; Sheen *Plant Cell* 2: 1027, 1990; or Jang and Sheen *Plant Cell* 6: 1665, 1994), and (7) the vortexing

- 27 -

method (see, e.g., Kindle *supra*). The method of transformation is not critical to the invention. Any method which provides for efficient transformation may be employed. As newer methods are available to transform crops or other host cells, they may be directly applied.

5 The following is an example outlining one particular technique, an *Agrobacterium*-mediated plant transformation. By this technique, the general process for manipulating genes to be transferred into the genome of plant cells is carried out in two phases. First, cloning and DNA modification steps are carried out in *E. coli*, and the plasmid containing the gene construct of interest is transferred by conjugation or
10 electroporation into *Agrobacterium*. Second, the resulting *Agrobacterium* strain is used to transform plant cells. Thus, for the generalized plant expression vector, the plasmid contains an origin of replication that allows it to replicate in *Agrobacterium* and a high copy number origin of replication functional in *E. coli*. This permits facile production and testing of transgenes in *E. coli* prior to transfer to *Agrobacterium* for
15 subsequent introduction into plants. Resistance genes can be carried on the vector, one for selection in bacteria, for example, streptomycin, and another that will function in plants, for example, a gene encoding kanamycin resistance or herbicide resistance. Also present on the vector are restriction endonuclease sites for the addition of one or more transgenes and directional T-DNA border sequences which, when recognized by
20 the transfer functions of *Agrobacterium*, delimit the DNA region that will be transferred to the plant.

 In another example, plant cells may be transformed by shooting into the cell tungsten microprojectiles on which cloned DNA is precipitated. In the Biolistic Apparatus (Bio-Rad) used for the shooting, a gunpowder charge (22 caliber Power
25 Piston Tool Charge) or an air-driven blast drives a plastic macroprojectile through a gun barrel. An aliquot of a suspension of tungsten particles on which DNA has been precipitated is placed on the front of the plastic macroprojectile. The latter is fired at an acrylic stopping plate that has a hole through it that is too small for the

- 28 -

macroprojectile to pass through. As a result, the plastic macroprojectile smashes against the stopping plate, and the tungsten microprojectiles continue toward their target through the hole in the plate. For the instant invention the target can be any plant cell, tissue, seed, or embryo. The DNA introduced into the cell on the microprojectiles becomes integrated into either the nucleus or the chloroplast.

In general, transfer and expression of transgenes in plant cells are now routine practices to those skilled in the art, and have become major tools to carry out gene expression studies in plants and to produce improved plant varieties of agricultural or commercial interest.

Transgenic Plant Regeneration

Plant cells transformed with a plant expression vector can be regenerated, for example, from single cells, callus tissue, or leaf discs according to standard plant tissue culture techniques. It is well known in the art that various cells, tissues, and organs from almost any plant can be successfully cultured to regenerate an entire plant; such techniques are described, e.g., in Vasil *supra*; Green et al., *supra*; Weissbach and Weissbach, *supra*; and Gelvin et al., *supra*.

In one particular example, a cloned C-14 sterol reductase polypeptide or an antisense construct under the control of the 35S CaMV promoter and the nopaline synthase terminator and carrying a selectable marker (for example, kanamycin resistance) is transformed into *Agrobacterium*. Transformation of leaf discs (for example, of tobacco leaf discs), with vector-containing *Agrobacterium* is carried out as described by Horsch et al. (*Science* 227: 1229, 1985). Putative transformants are selected after a few weeks (for example, 3 to 5 weeks) on plant tissue culture media containing kanamycin (e.g. 100 µg/ml). Kanamycin-resistant shoots are then placed on plant tissue culture media without hormones for root initiation. Kanamycin-resistant plants are then selected for greenhouse growth. If desired, seeds from self-fertilized transgenic plants can then be sowed in a soil-less medium and grown in a

greenhouse. Kanamycin-resistant progeny are selected by sowing surfaced sterilized seeds on hormone-free kanamycin-containing media. Analysis for the integration of the transgene is accomplished by standard techniques (see, for example, Ausubel et al. *supra*; Gelvin et al. *supra*).

5 Transgenic plants expressing the selectable marker are then screened for transmission of the transgene DNA by standard immunoblot and DNA detection techniques. Each positive transgenic plant and its transgenic progeny are unique in comparison to other transgenic plants established with the same transgene. Integration of the transgene DNA into the plant genomic DNA is in most cases random, and the
10 site of integration can profoundly affect the levels and the tissue and developmental patterns of transgene expression. Consequently, a number of transgenic lines are usually screened for each transgene to identify and select plants with the most appropriate expression profiles.

Transgenic lines are evaluated for levels of transgene expression. Expression at
15 the RNA level is determined initially to identify and quantitate expression-positive plants. Standard techniques for RNA analysis are employed and include PCR amplification assays using oligonucleotide primers designed to amplify only transgene RNA templates and solution hybridization assays using transgene-specific probes (see, e.g., Ausubel et al., *supra*). The RNA-positive plants are then analyzed for protein
20 expression by Western immunoblot analysis using C-14 sterol reductase specific antibodies (see, e.g., Ausubel et al., *supra*). In addition, *in situ* hybridization and immunocytochemistry according to standard protocols can be done using transgene-specific nucleotide probes and antibodies, respectively, to localize sites of expression within transgenic tissue.

25 Once the recombinant C-14 sterol reductase protein is expressed in any cell or in a transgenic plant (for example, as described above), it may be isolated, e.g., using affinity chromatography. In one example, an anti-C14 sterol reductase antibody (e.g., produced as described in Ausubel et al., *supra*, or by any standard technique) may be

- 30 -

attached to a column and used to isolate the polypeptide. Lysis and fractionation of C-14 sterol reductase-producing cells prior to affinity chromatography may be performed by standard methods (see, e.g., Ausubel et al., *supra*). Once isolated, the recombinant protein can, if desired, be further purified, for example, by high performance liquid chromatography (see, e.g., Fisher, *Laboratory Techniques In Biochemistry And Molecular Biology*, eds., Work and Burdon, Elsevier, 1980).

Polypeptides of the invention, particularly short C-14 sterol reductase protein fragments, can also be produced by chemical synthesis (e.g., by the methods described in *Solid Phase Peptide Synthesis*, 2nd ed., 1984 The Pierce Chemical Co., Rockford, IL).

These general techniques of polypeptide expression and purification can also be used to produce and isolate useful C-14 sterol reductase fragments or analogs.

Antibodies

C-14 sterol reductases described herein (or immunogenic fragments or analogs) may be used to raise antibodies useful in the invention; such polypeptides may be produced by recombinant or peptide synthetic techniques (see, e.g., *Solid Phase Peptide Synthesis*, 2nd ed., 1984, Pierce Chemical Co., Rockford, IL; Ausubel et al., *supra*). The peptides may be coupled to a carrier protein, such as KLH as described in Ausubel et al, *supra*. The KLH-peptide is mixed with Freund's adjuvant and injected into guinea pigs, rats, or preferably rabbits. Antibodies may be purified by peptide antigen affinity chromatography.

Monoclonal antibodies may be prepared using the C-14 sterol reductase polypeptides described above and standard hybridoma technology (see, e.g., Kohler et al., *Nature* 256:495, 1975; Kohler et al., *Eur. J. Immunol.* 6:511, 1976; Kohler et al., *Eur. J. Immunol.* 6:292, 1976; Hammerling et al., *In Monoclonal Antibodies and T Cell Hybridomas*, Elsevier, NY, 1981; Ausubel et al., *supra*).

- 31 -

Once produced, polyclonal or monoclonal antibodies are tested for specific C-14 sterol reductase recognition by Western blot or immunoprecipitation analysis (by the methods described in Ausubel et al., *supra*). Antibodies which specifically recognize C-14 sterol reductases are considered to be useful in the invention; such antibodies may be used, e.g., in an immunoassay to monitor the level of C-14 sterol reductase produced by a plant.

Use

Because the present invention provides for the genetic manipulation of a plant sterol biosynthetic pathway, this invention described is useful for a variety of agricultural and commercial purposes including, but not limited to, increasing crop yields, improving crop and ornamental quality, and reducing agricultural production costs. For example, the methods, DNA constructs, proteins, and transgenic plants described herein are useful for improving a number of fruit and vegetable characteristics including, but not limited to, texture, size, nutritional content, modification of sterol composition, disease and insect resistance, and ripening processes. In addition, genetic manipulation of plant sterol composition (for example, seed sterol composition) is useful for improving food quality and oil stability, and regulating the formation of compounds having anti-nutritional properties.

In one particular example, antisense C-14 sterol reductase sequences are useful for reducing the expression of C-14 sterol reductase expression in a transgenic plant. Such reduced expression of C-14 sterol reductase provides a means for increasing the life-span of such plants. Increased life-span extends reproductive period, delays senescence, and increases branch number for high productivity and yield. In addition, transgenic plants expressing antisense C-14 sterol reductase are useful for producing plants having reduced and more compact proportions. Such plants require less space and land requirements for their growth, and are more convenient and efficient to harvest.

- 32 -

Overproduction of the C-14 sterol reductase in transgenic plants is useful for enhancing the production of steroid compounds having a variety of medicinal or agricultural applications. For example, overproduction of mammalian steroid hormones in plants offers an inexpensive means for producing such hormones.

5 In addition, C-14 sterol reductase polypeptides disclosed herein are useful for the development of enzyme inhibitors of the sterol biosynthetic pathway.

Other Embodiments

10 In other embodiments, the invention includes any protein which is substantially identical to a crucifer C-14 sterol reductase polypeptide (Fig. 10; SEQ ID NO:1); such homologs include other substantially pure naturally-occurring plant C-14 sterol reductase proteins as well as allelic variants; natural mutants; induced mutants; proteins encoded by DNA that hybridizes to the C-14 sterol reductase DNA sequence of Fig. 14 (SEQ ID NO: 2) under high stringency conditions or, less preferably, under
15 low stringency conditions (e.g., washing at 2X SSC at 37°C with a probe length of at least 10-15 nucleotides), both as described herein; and proteins specifically bound by antisera directed to a C-14 sterol reductase polypeptide. The term also includes chimeric polypeptides that include a C-14 sterol reductase portion.

The invention further includes analogs of any naturally-occurring plant C-14
20 sterol reductase polypeptide. Analogs can differ from the naturally-occurring C-14 sterol reductase protein by amino acid sequence differences, by post-translational modifications, or by both. Analogs of the invention will generally exhibit at least 30%, more preferably 40%, and most preferably 50% or even 80-95% identity with all or part of a naturally-occurring plant C-14 sterol reductase amino acid sequence. The
25 length of sequence comparison is at least 15 amino acid residues, preferably at least 25 amino acid residues, and more preferably more than 35 amino acid residues. Modifications include *in vivo* and *in vitro* chemical derivatization of polypeptides, e.g., acetylation, carboxylation, phosphorylation, or glycosylation; such modifications

- 33 -

may occur during polypeptide synthesis or processing or following treatment with isolated modifying enzymes. Analogs can also differ from the naturally-occurring C-14 sterol reductase polypeptide by alterations in primary sequence. These include genetic variants, both natural and induced (for example, resulting from random mutagenesis by irradiation or exposure to ethanemethylsulfate or by site-specific
5 mutagenesis as described in Sambrook, Fritsch and Maniatis, *Molecular Cloning: A Laboratory Manual* (2d ed.), CSH Press, 1989, or Ausubel et al., *supra*). Also included are cyclized peptides, molecules, and analogs which contain residues other than L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino
10 acids, e.g., β or γ amino acids.

In addition to full-length polypeptides, the invention also includes C-14 sterol reductase polypeptide fragments. As used herein, the term "fragment," means at least 20 contiguous amino acids, preferably at least 30 contiguous amino acids, more preferably at least 50 contiguous amino acids, and most preferably at least 60 to 80 or
15 more contiguous amino acids. Fragments of C-14 sterol reductase polypeptides can be generated by methods known to those skilled in the art or may result from normal protein processing (e.g., removal of amino acids from the nascent polypeptide that are not required for biological activity or removal of amino acids by alternative mRNA splicing or alternative protein processing events).

Furthermore, the invention includes nucleotide sequences that facilitate specific detection of a C-14 sterol reductase nucleic acid. Thus, C-14 sterol reductase sequences described herein (e.g., SEQ ID NO: 2 and 3) or portions thereof may be used as probes to hybridize to nucleotide sequences from other plants (e.g., dicots, monocots, gymnosperms, and algae) by standard hybridization techniques under
20 conventional conditions. Sequences that hybridize to a C-14 sterol reductase coding sequence or its complement and that encode a C-14 sterol reductase are considered useful in the invention. As used herein, the term "fragment," as applied to nucleic acid sequences, means at least 5 contiguous nucleotides, preferably at least 10 contiguous
25

- 34 -

nucleotides, more preferably at least 20 to 30 contiguous nucleotides, and most preferably at least 40 to 80 or more contiguous nucleotides. Fragments of C-14 sterol reductase nucleic acid sequences can be generated by methods known to those skilled in the art.

5 All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication or patent application was specifically and individually indicated to be incorporated by reference.

Other embodiments are within the following claims.

10

What is claimed is:

Claims

1. A substantially pure plant C-14 sterol reductase polypeptide.
- 5 2. A substantially pure polypeptide comprising an amino acid sequence substantially identical to the sequence shown in Fig. 14 (SEQ ID NO: 1).
3. The polypeptide of claim 1 or 2, wherein said polypeptide comprises the amino acid sequence shown in Fig. 14 (SEQ ID NO: 1).
- 10 4. The polypeptide of claim 1 or 2, wherein said polypeptide is from a dicot.
5. The polypeptide of claim 4, wherein said dicot is a crucifer.
- 15 6. A purified DNA encoding a C-14 sterol reductase polypeptide.
7. A purified DNA comprising a sequence substantially identical to the DNA sequence shown in Fig. 14 (SEQ ID NO: 2).
- 20 8. The purified DNA of claim 6 or 7, wherein said DNA comprises the sequence shown in Fig. 14 (SEQ ID NO: 2).
9. The purified DNA of claim 6 or 7, wherein said DNA encodes a polypeptide which has an amino acid sequence substantially identical to that shown in Fig. 14
25 (SEQ ID NO: 1).
10. The purified DNA of claim 9, wherein said DNA encodes a polypeptide which has the amino acid sequence shown in Fig. 14 (SEQ ID NO: 1).

- 36 -

11. The purified DNA of claim 6 or 7, wherein said DNA is from a dicot.

12. The purified DNA of claim 11, wherein said dicot is a crucifer.

5 13. A vector comprising the purified DNA of claim 6 or 7.

14. A cell comprising the purified DNA of claim 6 or 7.

10 15. A method of producing a recombinant C-14 sterol reductase polypeptide
comprising
providing a cell transformed with purified DNA encoding a C-14 sterol
reductase polypeptide positioned for expression in said cell, and
culturing said transformed cell under conditions for expressing said DNA.

15 16. The method of claim 15, wherein said method further comprises recovering
said recombinant C-14 sterol reductase polypeptide.

17. The method of claim 15, wherein said cell is a plant cell.

20 18. A recombinant C-14 sterol reductase produced by the method of claim 15.

19. An isolated antibody which specifically recognizes and binds a plant C-14
sterol reductase polypeptide.

25 20. A transgenic plant which contains DNA encoding a C-14 sterol reductase
polypeptide integrated into the genome of said plant, wherein said DNA is expressed
in said transgenic plant.

- 37 -

21. A transgenic plant which contains DNA encoding an amino acid sequence substantially identical to the sequence shown in Fig. 14 (SEQ ID NO: 1) integrated into the genome of said plant, wherein said DNA is expressed in said transgenic plant.

5

22. A seed from a transgenic plant of claim 20 or 21.

23. A cell from a transgenic plant of claim 20 or 21.

10

24. A method of detecting a C-14 sterol reductase gene in a plant cell comprising:

contacting the purified DNA of claim 6 or 7 or a portion thereof greater than about 12 nucleotides in length with a preparation of genomic DNA from said plant cell under hybridization conditions providing detection of DNA sequences having about

15

40% or greater sequence identity to SEQ ID NO: 2.

25. A method of isolating a C-14 sterol reductase gene or portion thereof, said method comprising

20

(a) amplifying said C-14 sterol reductase gene using oligonucleotide primers, wherein said primers each have regions of complementarity to opposite DNA strands in a region of SEQ ID NO: 2; and

(b) isolating said C-14 sterol reductase gene or portion thereof.

25

26. A method for reducing the level of a C-14 sterol reductase polypeptide in a transgenic plant cell, said method comprising expressing in a plant cell an antisense C-14 sterol reductase nucleic acid sequence.

- 38 -

27. The method of claim 26, wherein said antisense C-14 sterol reductase nucleic acid sequence is encoded by a transgene integrated into the genome of said transgenic plant cell.

5 28. The method of claim 26, wherein said C-14 sterol reductase sequence is SEQ ID NO: 2 or SEQ ID NO: 3.

 29. The method of claim 26, wherein said method further comprises growing a transgenic plant from said transgenic plant cell, whereby the level of the C-14 sterol
10 reductase polypeptide is reduced in said transgenic plant.

 30. A method for increasing the level of a C-14 sterol reductase in a transgenic plant cell, comprising expressing in said transgenic plant cell a nucleic acid sequence encoding a polypeptide substantially identical to SEQ ID NO: 1.
15

 31. A transgenic plant comprising purified DNA encoding a plant C-14 sterol reductase polypeptide, said DNA comprising a knockout mutation in said C-14 sterol reductase sequence.



Fig. 1A



Fig. 1B

SUBSTITUTE SHEET (RULE 26)

2/29

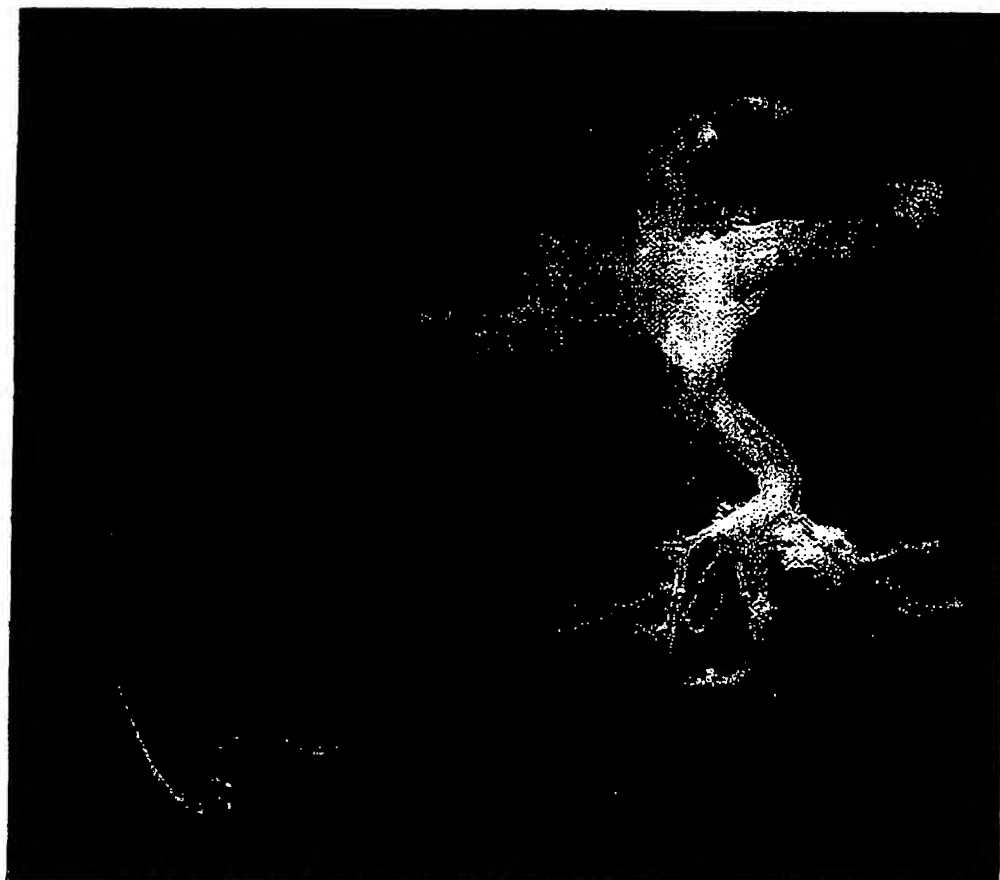


Fig. 2

SUBSTITUTE SHEET (RULE 26)

3/29



Fig. 3A



Fig. 3B

SUBSTITUTE SHEET (RULE 26)

4/29



Fig. 4

SUBSTITUTE SHEET (RULE 26)

5/29



Fig. 5

SUBSTITUTE SHEET (RULE 26)

6/29

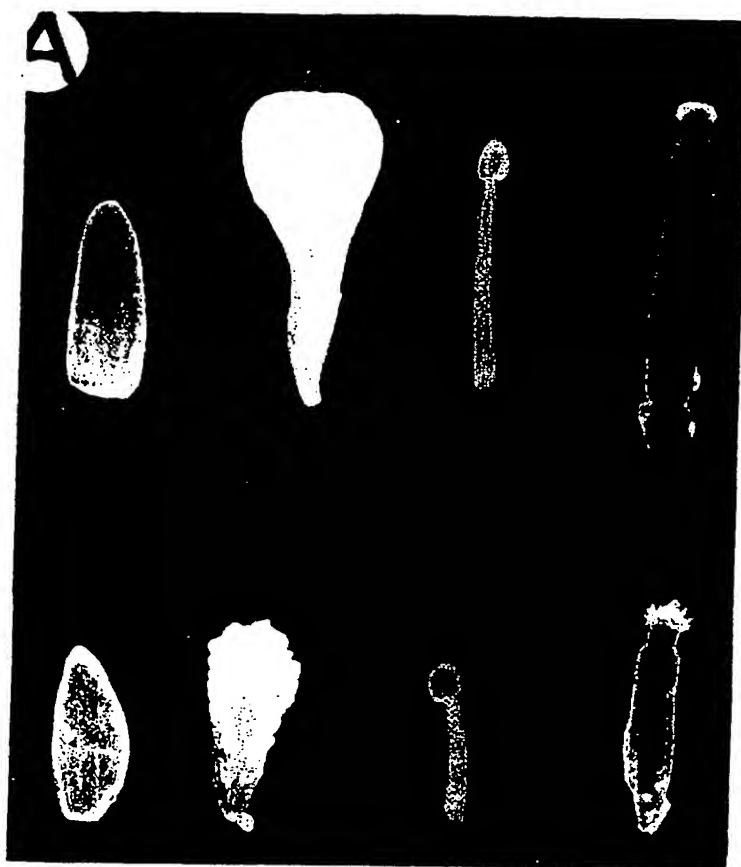


Fig. 6A

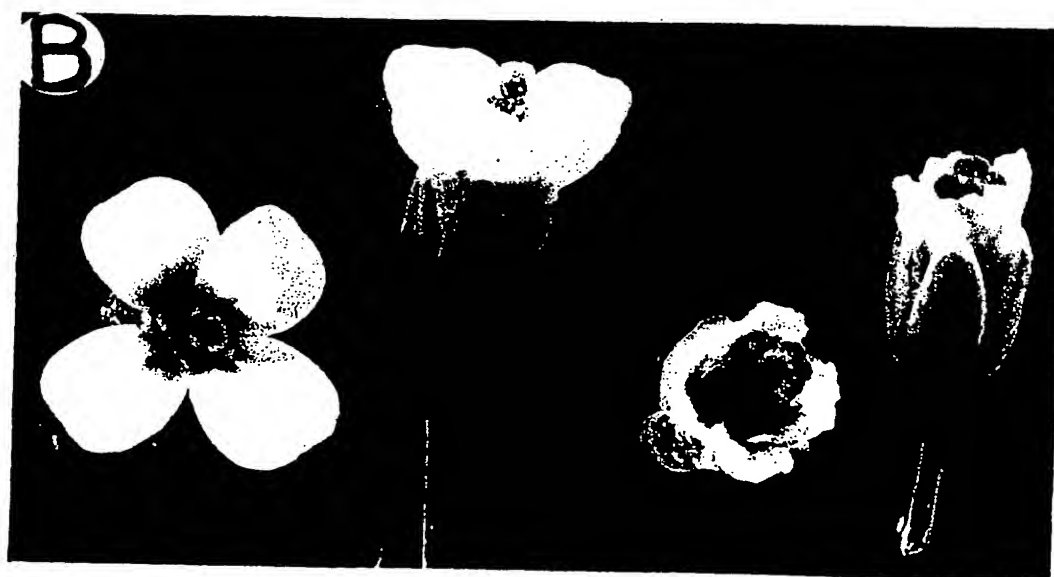


Fig. 6B

SUBSTITUTE SHEET (RULE 26)

7/29

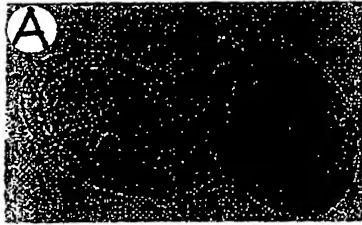


Fig. 7A

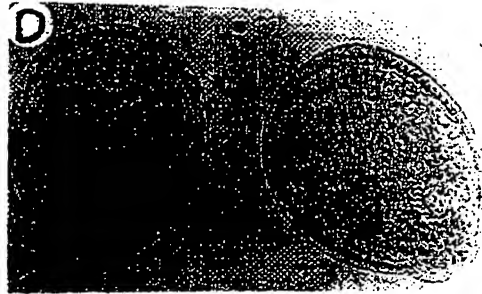


Fig. 7D



Fig. 7B

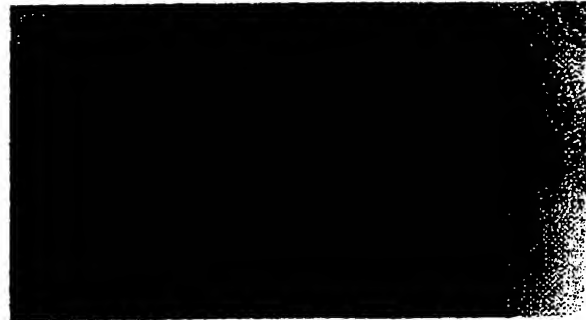


Fig. 7E

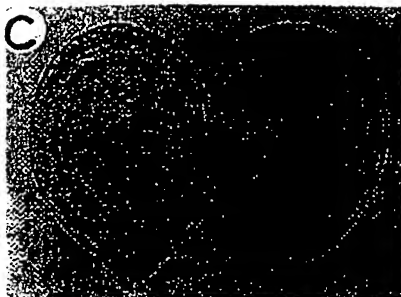


Fig. 7C



Fig. 7F

SUBSTITUTE SHEET (RULE 26)

8/29

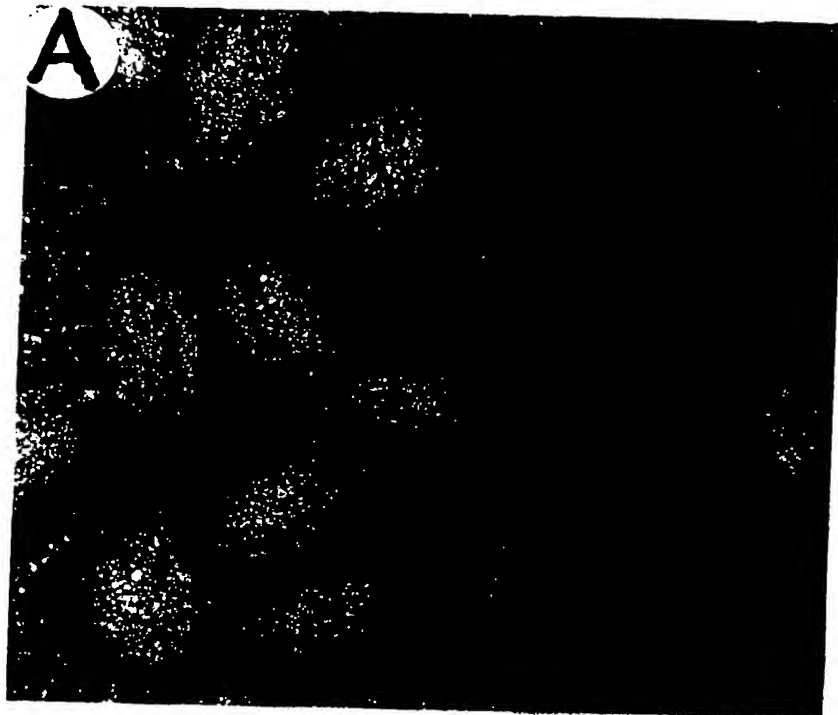


Fig. 8A

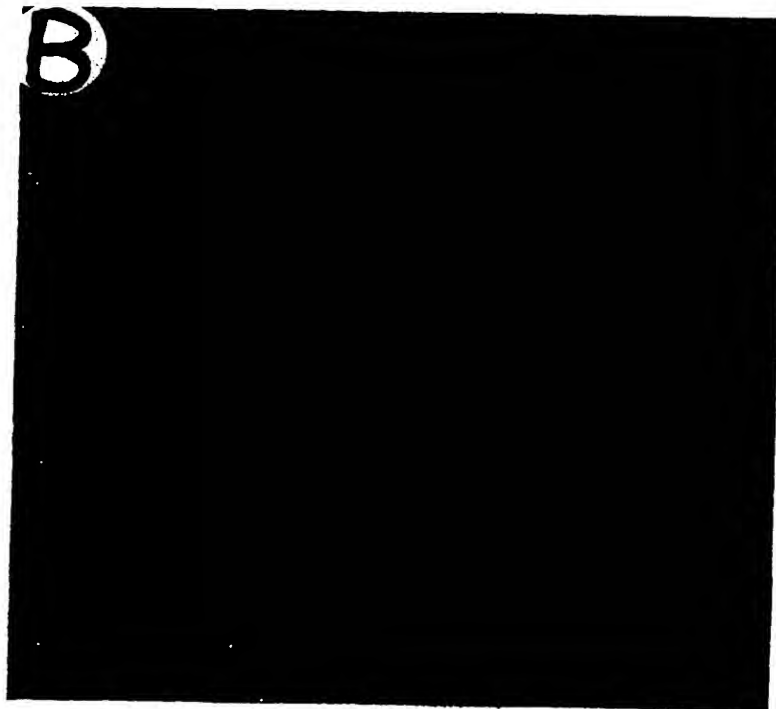


Fig. 8B

SUBSTITUTE SHEET (RULE 26)

9/29



Fig. 9A



Fig. 9D

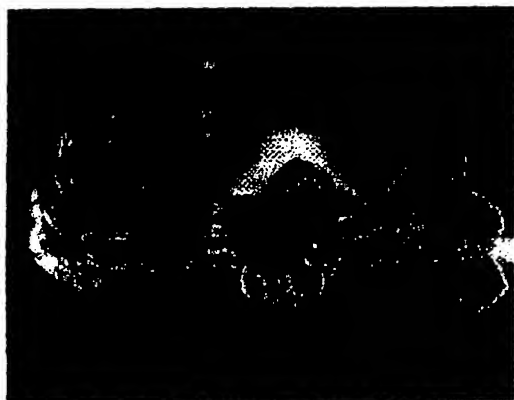


Fig. 9B



Fig. 9E



Fig. 9C



Fig. 9F

SUBSTITUTE SHEET (RULE 26)

10/29

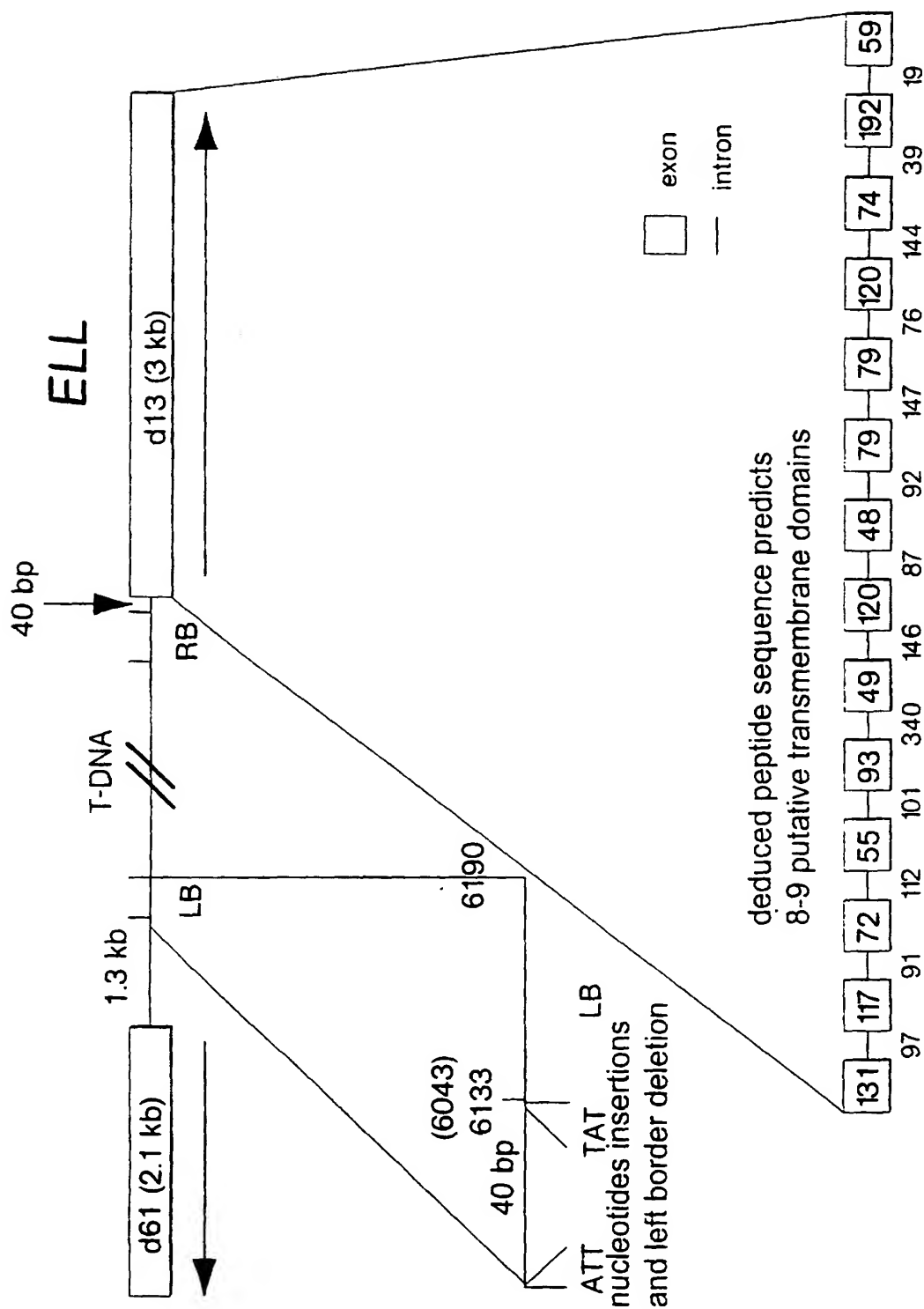


Fig. 10A

SUBSTITUTE SHEET (RULE 26)

11/29

chr3

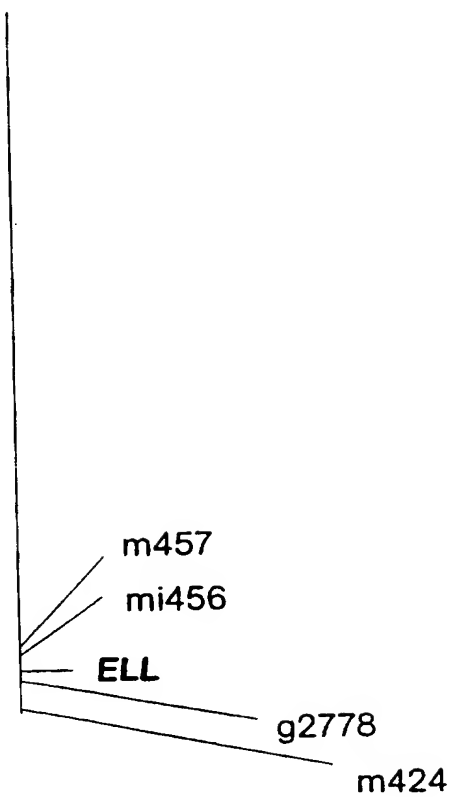


Fig. 10B

SUBSTITUTE SHEET (RULE 26)

12/29

SEQ ID NO: 1	Ath	0
SEQ ID NO: 4	Erg24	19
SEQ ID NO: 5	Pombe	20
SEQ ID NO: 6	Sts1	18
SEQ ID NO: 7	Yg1022	MAKDNSEKQLQ	VOGEEKKSKQ	PVNFLPQGW	LKPNEI	45
Ath	14
Erg24	60
Pombe	63
Sts1	62
Yg1022	87
Ath	QSVYVLV...	46
Erg24	102
Pombe	95
Sts1	107
Yg1022	132
Ath	87
Erg24	146
Pombe	138
Sts1	149
Yg1022	174
Ath	123
Erg24	191
Pombe	172
Sts1	185
Yg1022	206
Ath	166
Erg24	235
Pombe	216
Sts1	230
Yg1022	250
Ath	209
Erg24	279
Pombe	261
Sts1	274
Yg1022	294
Ath	251
Erg24	321
Pombe	303
Sts1	319
Yg1022	339
Ath	283
Erg24	349
Pombe	335
Sts1	364
Yg1022	384
Ath	325
Erg24	394
Pombe	380
Sts1	409
Yg1022	429
Ath	369
Erg24	438
Pombe	424
Sts1	453
Yg1022	473

Fig. 11

SUBSTITUTE SHEET (RULE 26)

13/29

SEQ ID NO: 1	Ath	0
SEQ ID NO: 8	Chicken	MPNRYVADGE	VVGRWEGSV	LYYEVQVTSY	DDASHLYTVK	40
SEQ ID NO: 9	Human	MPSRKVADGE	VVRGRWPGSS	LYYEVVILSH	DSTSLYTVK	40
	Ath	0
	Chicken	YKDGTELAALK	ESDIRLQSSF	KORKSQSSSS	SPSRR..SRS	78
	Human	YKDGTELELK	ENDIKPLTSF	RQRKGGSTSS	SPSRRRGSR	80
	Ath	0
	Chicken	RSRSRSPGRP	AKGRRRSSSH	SREHKEDKKK	IIQETSTAP	117
	Human	RSRSRSPGRP	EKSARRSASA	SHQADIKEAR	REVEVKLTPL	120
	Ath	0
	Chicken	.PKPSENTR	RYNGEPDSTE	RNDTSSKLL	QOKLPDVEM	156
	Human	ILKPFGNIS	RYNGEPHIE	RNDAPHK.NT	QOKFNLSQES	159
	Ath	0
	Chicken	ERVLDQYSLR	SRREEKK...	...KEETVAAE	KKIFEALKTP	190
	Human	SYIATQYSLR	FRREEVKLKE	IDSKEEKYVA	KEL...AVRTF	197
	Ath	4
	Chicken	EKPSSKTKEL	EFGGRFGTEM	LMFELPATVL	YLVLCKODD	230
	Human	EVTPIRAKDL	EFGGVPGVEL	IMFGLPVFLF	ILLVLCKOKD	237
	Ath	...MDLGVLT	PSLOSVM...	...VLVELEV	YIAVAGEILP	35
	Chicken	PSLMNF.PPL	PALESILWETK	VFGVELLWFF	FOALEYLLPI	269
	Human	PSLENEPPL	PALVELWETR	VEGVYLLWEL	IQVLEYLLPI	277
	Ath	GKVERGVLLS	DGSOLRYR.N	GILALILLVA	ILGICAKLGI	75
	Chicken	GKVVEGLPLS	NERKLOYRIN	GEYAFLLTAA	AMGTLLEYFQ	309
	Human	GKVVEGTPLI	DGRRLKYRIN	GEYAFILTS	VIGTSLFQGV	317
	Ath	VSPVAVADRG	LEELSAFFIF	CVLVTLALYV	TGRSSSNKGS	115
	Chicken	ELHYLY.DHE	VQFAVSAAF	SMALSYLYI	..RSLKAPPE	346
	Human	EHHVYV.SHE	LOFALAATVF	CVVLSVLYLM	..RSLKAPRN	354
	Ath	SLRPHV.SGN	LVHDWWEGLQ	LNPQFMSIDL	KEFF.VRAGM	153
	Chicken	DLAPGGNSGY	LVYDFETGHE	LNPRIGSFDL	KYFCELRPGL	386
	Human	DLSPAS.SGN	AVYDFETGRE	LNPRIGTFDL	KYFCELRPGL	393
	Ath	MGWLLINLSI	L..AKSVODG	SL.SOSMILY	QIFCALYELD	190
	Chicken	IGWVVINLAM	LLAEMKIHNQ	SMPSLSMILV	NSFQLLYVVD	426
	Human	IGWVVINLVM	LLAEMKIQDR	AVPSLAMILV	NSFQLLYVVD	433
	Ath	YFVHEEYMTS	TWDILAEREG	FMLVFGDLW	IPFTFSLQGW	230
	Chicken	ALWNEEAULT	TMDIITHDGFG	FMLAFGDLVW	VPFVYSLQAF	466
	Human	ALWNEEAULT	TMDIITHDGFG	FMLAFGDLVW	VPFVYSLQAF	473
	Ath	WELHNKVELT	VPAIVVNCLV	FLLGYMVERG	ANKOKHIFKK	270
	Chicken	YLVGHPIATS	WPVAAATIL	NCIGYIFRS	ANSOKNFRR	506
	Human	YLVSHPEVS	WPMASLILV	KLCGYVIFRG	ANSOKNAFRK	513
	Ath	NPKTPIWGKP	PVV...VGCK	LLVSGYWGIA	RHONYLGOLM	307
	Chicken	NPADPKLSYL	KVIFTATGKG	LLVIGWGWGFV	RHPNYLGDII	546
	Human	NPSDPKLAHL	KTIHTSTCKN	LLVSGWGWGFV	RHPNYLGDII	553
	Ath	BALSFSLPCG	ISSPVYFYF	IYLLILLIAR	ERRDEVRCAE	347
	Chicken	MALAWSLPCG	FNHILPYFYV	IYFICLLVHR	EARDEHCKK	586
	Human	MALAWSLPCG	FNHILPYFYI	IYFTMLLVHR	EARDEVHCKK	593
	Ath	KYKEIWAAYL	RLVPRILPY	VY.....	369
	Chicken	KYGLAWERYC	QRVPYTHISL	HMLEHSTYLI	CKLKYTSHLC	626
	Human	KYGVAWERYC	QRVPYRIFFY	IY.....	615
	Ath369
	Chicken	TWSVCYLGRK	H637	H637
	Human615

Fig. 12

SUBSTITUTE SHEET (RULE 26)

14/29

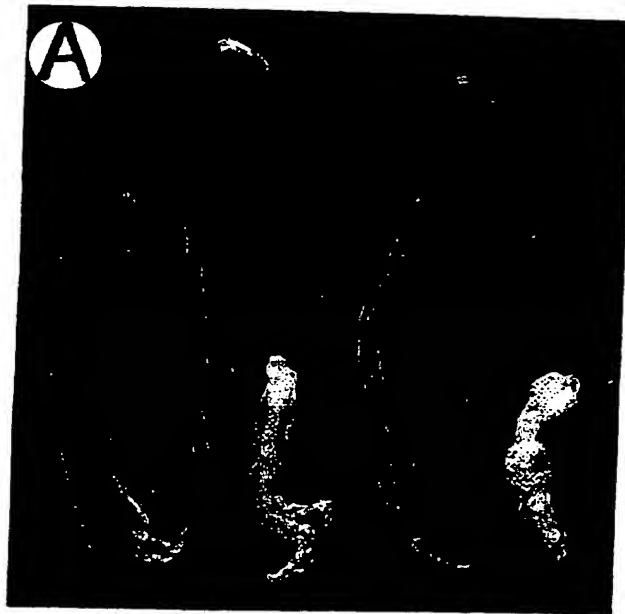


Fig. 13A



Fig. 13B

SUBSTITUTE SHEET (RULE 26)

15/29

SEQ ID NO:2

CTGAAATTAAACAAAGCGAGAAAAGGCGATACAAACGATTTTGAATGCTTCATCTTCTCC

1 -----+-----+-----+-----+-----+-----+ 60
GACTTTAATTTGTTTCGCTCTTTTCCGCTATGTTTGCTAAAGCTTACGAAGTAGAAGAGG

c

TTTGAAAATCCTTCTTCTGCTTAATGCTGCTAGATATGGATCTCGGTGTTCTTCTTCCAT

61 -----+-----+-----+-----+-----+-----+ 120
AAACTTTTAGGAAGAAGACGAATTACGACGATCTATACCTAGAGCCACAAGAAGAAGGTA

c

SEQ ID NO: 1 M L L D M D L G V L L P S -

CATTGCAATCTGTTTATGTGCTGGTGTGTTTACTTTCGTTTACTTGGCCGTTGCCGGAGAAA

121 -----+-----+-----+-----+-----+-----+ 180
GTAACGTTAGACAAATACACGACCACAAAATGAAGCAAATGAACCGGCAACGGCCTCTTT

c

L Q S V Y V L V F Y F V Y L A V A G E I -

TTCTCCCCGGGAAAGTTATTCGCGGCGTCCTTTTATCAGATGGCTCTCAACTTCGTTACC

181 -----+-----+-----+-----+-----+-----+ 240
AAGAGGGGCCCTTTCAATAAGCGCCGCAGGAAAATAGTCTACCGAGAGTTGAAGCAATGG

c

L P G K V I R G V L L S D G S Q L R Y R -

Fig. 14-1

SUBSTITUTE SHEET (RULE 26)

16/29

GATGCAATGGTCTATTGGCACTAATATTGTTGGTAGCTATTTTGGGAATCTGTGCAAAAC
241 -----+-----+-----+-----+-----+-----+ 300
CTACGTTACCAGATAACCGTGATTATAACAACCATCGATAAAACCCTTAGACACGTTTTG
c C N G L L A L I L L V A I L G I C A K L -

TTGGCATTGTATCACCTCTTGTGGTTGCGGATAGAgGACTTGAgTTACTCTCAGCTACTT
301 -----+-----+-----+-----+-----+-----+ 360
AACCGTAACATAGTGGAGAACACCAACGCCTATCTcCTGAACtAATGAGAGTCGATGAA
c G I V S P L V V A D R G L E L L S A T F -

TTATTTTCTGTGTTTTGGTGACATTAgCATTGTATGTTACTGGGCGAAGTtCCTCgAATA
361 -----+-----+-----+-----+-----+-----+ 420
AATAAAAGACACAAAACCACTGTAATcGTAACATACAATGACCCGCTTCAaGGAGcTTAT
c I F C V L V T L A L Y V T G R S S S N K -

AggGtTctTCCCTAAAgCCTCATGTCtCAgGAAATCTTGTACATGACTGGTGGTTTGGAA
421 -----+-----+-----+-----+-----+-----+ 480
TccCaAgaAGGGATTTcGGAGTACAGAGTcCTTTAGAACATGTACTGACCACCAAACCTT
c G S S L K P H V S G N L V H D W W F G I -

Fig. 14-2

SUBSTITUTE SHEET (RULE 26)

17/29

TACAGCTGAATCCTCAGTTTATGAGCATTGATCTCAAGTTTTTCTTTGTCAGAGCCGGGA
481 -----+-----+-----+-----+-----+-----+ 540
ATGTCGACTTAGGAGTCAAATACTCGTAACTAGAGTTCAAAAAGAAACAGTCTCGGCCCT

c Q L N P Q F M S I D L K F F F V R A G M -

TGATGGGATGGCTGCTTATCAATCTCTCTATTCTGGCAAAAAGTGTGCAGGATGGTTCCT
541 -----+-----+-----+-----+-----+-----+ 600
ACTACCTTACCGACGAATAGTTAGAGAGATAAGACCGTTTTTCACACGTCCTACCAAGGA

c M G W L L I N L S I L A K S V Q D G S L -

TGAGTCAGTCGATGATTCTTTACCAGATCTTCTGTGCGTTATATATATTGGACTACTTTG
601 -----+-----+-----+-----+-----+-----+ 660
ACTCAGTCAGCTACTAAGAAATGGTCTAGAAGACACGCAATATATATAACCTGATGAAAC

c S Q S M I L Y Q I F C A L Y I L D Y F V -

TTTCATGAAGAATACATGACCTCTACGTGGGACATAATTGCAGAGAGACTAGGCTTCATGC
661 -----+-----+-----+-----+-----+-----+ 720
AAGTACTTCTTATGTACTGGAGATGCACCCTGTATTAACTCTCTCTGATCCGAAGTACG

c H E E Y M T S T W D I I A E R L G F M L -

Fig. 14-3

SUBSTITUTE SHEET (RULE 26)

18/29

721 TAGTGTTTGGAGATCTCCTGTGGATTCTTTTCACTTTTAGCATTTCAGGGCTGGTGGCTTT
-----+-----+-----+-----+-----+-----+ 780
ATCACAAACCTCTAGAGGACACCTAAGGAAAGTGAAAATCGTAAGTCCCACCACCGAAA
C V F G D L L W I P F T F S I Q G W W L L -

781 TGCACAACAAAGTAgAACTAACAGTTCCTGCGATTGTAGTCAATTGCCTTGTCTTCTTGA
-----+-----+-----+-----+-----+-----+ 840
ACGTGTTGTTTCATcTTGATTGTCAAGGACGCTAACATCAGTTAACGGAACAGAAGAACT
C H N K V E L T V P A I V V N C L V F L I -

841 TAGGGTACATGGTTTTTTCGAgGAgCTAACAAACAAAAACATATCTTTAAGAAGAACCCAA
-----+-----+-----+-----+-----+-----+ 900
ATCCCATGTACCAAAAAGCTcCTcGATTGTTTGTGTTTTGTATAGAAATTCCTTCTGGGTT
C G Y M V F R G A N K Q K H I F K K N P K -

901 AAACACCAATATGGGGCAAGCCTCCAGTGGTAGTTGGTGGAAAGTTACTGGTTCAGGCT
-----+-----+-----+-----+-----+-----+ 960
TTTGTGGTTATACCCCGTTCGGAGGTCACCATCAACCACCTTTCAATGACCAAAGTCCGA
C T P I W G K P P V V V G G K L L V S G Y -

Fig. 14-4
SUBSTITUTE SHEET (RULE 28)

19/29

ATTGGGGAATTGCAAGGCACTGTAATTACCTTGGCGACTTGATGCTTGCTCTGTCCTTCA
961 -----+-----+-----+-----+-----+-----+ 1020
TAACCCCTTAACGTTCCGTGACATTAATGGAACCGCTGAACTACGAACGAGACAGGAAGT

c W G I A R H C N Y L G D L M L A L S F S -

GTTTGCCATGTGGAATAAGTTCTCCGGTTCCATATTTCTACCCGATATACTTCTGATAC
1021 -----+-----+-----+-----+-----+-----+ 1080
CAAACGGTACACCTTATTCAAGAGGCCAAGGTATAAAGATGGGCTATATGGAAGACTATG

c L P C G I S S P V P Y F Y P I Y L L I L -

TATTGATATGGAGAGAACGAAGAGACGAGGTTTCGATGTGCAGAGAAGTACAAGGAGATAT
1081 -----+-----+-----+-----+-----+-----+ 1140
ATAACTATACCTCTCTTGCTTCTCTGCTCCAAGCTACACGTCTCTTCATGTTCTCTCTATA

c L I W R E R R D E V R C A E K Y K E I W -

GGGCAGAGTATCTTAGACTTGTCCTTGGGAGAACTTCCTTATGTTTATTAGATGTGCC
1141 -----+-----+-----+-----+-----+-----+ 1200
CCCGTCTCATAGAATCTGAACAGGGGACCTCTTATGAAGGAATACAAATAATCTACACGG

c A E Y L R L V P W R I L P Y V Y *

Fig. 14-5
SUBSTITUTE SHEET (RULE 28)

20/29

AAGAGCCAAGTCATGAATCCTTTCAGATTCACCTCTGTTGTCTTATTTTTTCCATAATC
1201 -----+-----+-----+-----+-----+-----+ 1260
TTCTCGGTTTCAGTACTTaGGAAAGTCTAAGTGGAGAACAACAGAATAAAAAAGGTATTAG

c

TTGTTTTATTTTAGCAATGCTCGAATTGAACTTTGTAGTACACTTTTGAAAAATAACTT
1261 -----+-----+-----+-----+-----+-----+ 1320
AACAAAATAAAATCGTTACGAGCTTAACTTTGAAACATCATGTGAAACCTTTTATTGAA

c

CAGTCCTTAAAAAAAAAAAAAACCTAANTTACTCCCNCTGGGCGGCCGCTGGTTTTATAT
1321 -----+-----+-----+-----+-----+-----+ 1380
GTCAGGAATTTTTTTTTTTTTTTGGATTNAATGAGGGNGACCCGCCGCGACCAAAATATA

c

TTGTTGTAAAAATTAAANAATTACTNCCTTGANGATCTGTAAAAAAAAA
1381 -----+-----+-----+-----+-----+ 1429
AACACATTTTAAATTTNTTAATGANGGAACCTNCCTAGACATTTTTTTTTT

Fig. 14-6
SUBSTITUTE SHEET (RULE 26)

21/29

SEQ ID NO: 3

1 T N T T G A A G G N T N A A G A A A A A N T A N G G T A A G C T G G G N A G G A C A A G A N T T C T
51 T G T N A C C A C A A C A C A A C A A C G C C A T G A A C C N A T C G G T T T C T T N T G T T T N G
101 A G A T C A C C T T T C T T G A G T T G G T G G T T T C T G A G N T C A A G N T C C T T G T T G A C
151 T C A G T G A A G T C C A G A T G C A G C N T C A A A A C T T T T G T C C T G T A G A C N T A G C A
201 A G A G T A A C A G C A C C A A C C A A A T C G C T A T C C G A T G T A A T C A A A C C T T A T C
251 A C N T T C A T C G T C C T C A T A T A T A A T C T G A G G C C G T T G T T C C A C A T T G T T A T
301 T G T C G C T G C C A A T T C T T T G C A T C A C A A T A C C C A T C A G C T N T T C G A G G T T T
351 T C A G C T C C A G A A G T A A A C C G A T G T A C A C G G C C C T T A A G G T C T T C A A A T T T
401 G A A C G A A A A C G A A T T C C C T A G T C C T A G A G A T G G G T A A G A A C T G A G C T T C C
451 C T A T A T C T G A A T G A T G C A T C A T T G C C G A C A T T T C A C T T T G A G T G C A G A A
501 T C A T C A G G T G G C T C T A A C G C A A G A G C T G A A T C C C A A A A T T T C T G C A T C A T
551 C G T G T T T G C C A T A T C A T T T A C A G C T C C A G A A C T G T T C T C C A C C A T T G A A A
601 T A G C T G C G T G A G T A A T C T G A A G A A C G T C T A C A C A A G C T G C A G C T G A T C C A
651 T C T T T A T C T A T A A T T G G A A G A T G T A G A A A C T T T C C A T C A T G C A T T G T A T G
701 C A A T G C A T C C A G A A T C G T T G T C T C T A G C G A T G C A C A T T C A G G A T T C G G T G
751 T C A T T A C C T T C T C G A C A A G A G T C A A T T C A G G A G A T A A A T T T T G T G C C A C C
801 A C T C G C A T C A G A A T G T C C T T T G A A G T C A A G A T T C C A C T G A T T T T G T T C C C
851 C C G T G G A A A T G A T T A C A G A G T T A A C C C G C A A A T C C C T C A T C C T T T T C G C A
901 G C A A C T G A A A C A G G A T C T G A T G G T G C T A C A A G T G C A A C C T T C G A T G T C T G
951 T A A T A A T C G T G A C A A G G C G G G T T A A A C A T T N T N T C C T T C A A G G T T T C A A T
1001 G A A A G C A T A C G G T G C A G A A T A T C C G C T T C C C C A T T G T T T N T C C A C A C C T T
1051 C C A C T G C A G C A G C T A A A G C A C T A C C T T G C T C T G C A G T T T C T C C A T C C T A G
1101 A A A T A G C A T C A T A C A A A C A C T T T G T A A T A T C C A A C A A A G C A A T G A C T T C A
1151 C C A T T C T C C A C A A C A G G C A A G T G T C T A A A C T T C C C T T G A A C C A T C T T C T G
1201 A A G A G C C T C A A G C G C C A A C G A A T C A G A A G T A A C A A A A T A G G A T T C C T A G
1251 T C A T A A C C T T A G A G A C C A A A G T T T G A T C C G G T C T C A A C C C T T C A G C A A T C
1301 A C T C T T G T A G C T A C A T C T T T A T C A G T A A C A A T C C C G G A A A G A A G C G C A C T

Fig. 15-1
SUBSTITUTE SHEET (RULE 28)

22/29

SEQ ID NO: 3

1351 TGAATCAGTC AACAAACAAG CATCAACACG CCTAGCAGCC ATTNTTCGAC
1401 AAGCATCGAA AACAGTAGTT CCTTCNAGGA ATAGTAAGAG CTTTCGATAA
1451 CCTAAGCTTC TTCGCTGTCT CTCTCCATTA GAAGGAGCTT GAGATTGAGG
1501 TTGAGGAGGA GGTGAATTGG GTTTTGAGGT GTTCCCANTA ACACTTCCAT
1551 TCTCTGATTG TACTGGTTTC TTAGAAGGTG GTGGTCCTCN CCGTACAGTA
1601 GAATTGCTTC TCCTCCCTGA TGTGAAGAA GGACCCGTCG CTTGAGTACT
1651 CATATTCGGT CAATCTAGGG TTTACTTAGA TCCTAAATCC GTCANAAATG
1701 ATTCCTTTAG ATATCAAAC CGTCTCTGCA AATGAAAAAT TCAACCTTTA
1751 ATTCACAAAC TATTGAAATT TCATCTAAAG CACGAATCTG AATAAAACCC
1801 AATTCACAAT AAAGACGATT TGCTCTGAGA ATACGATGCA ACATACACGA
1851 AAAGGATTCG AATTAAACGG ACGAGGGAAA TGAAACAACT TGAAACCTA
1901 AGGATTTGAG CAGAAGTTAT GTGGGAAGAT TGGGNATTTA GGGTTTACCT
1951 TCTTCTTTCT TCNTCAAGGT CTCTCTCTCG AGCACTTTCTG TTNCCCCAAA
2001 AACNAACGGC TCTTAACAAT TGAGTTAANC CANTTATCGA GTTTTCATTG
2051 GNTGTTCCCTG TTTCCGCGTG TGTGGTGGNT CNCCACCTCC TTTCTTATAA
2101 TCNACGACTA AAAATGTTAA ANATAANACT AANATTTCTT TCTANAAAAA
2151 TCGTAAAANC CAAATGTTTT TTTTTTTCTG ATAAATGTCT ATAAATCACC
2201 CTTTCTTTTT AAATAATGAA ATTTGATGAC ATTTATCTCT TGTATCTAGN
2251 AGAGTTAATG GCTAACATAA ANACCAGAAA AAATTAATTC NAATAAATAT
2301 GATTTGTGTG GGTTACATGG AAAAATTGTC AAATAATAAA NCAAAAAAAA
2351 ATTGTATAGA TGCAGTGCAA GTTGTTTCTG GTCAACTTGC CGTCGAGCCT
2401 CACAACTGTT TGTTACAAGT GGACTCGCAT GTAATTCCTT CTTTTAATAA
2451 CTTACCAGTT ACACCATCCA ACATGTGATT TGACAGAAAA ATATTTTAGT
2501 GAAATGTGAT CGGTGCAGAT TTTTCTATGT ACGTTTAAGC CTTTAAGGTA
2551 GACGTTTAAT CCNAAAATAT CCCTGAATAA CAACACCGAT TAATGGAACC
2601 AAGTAGATAC CTCCTCCGTT TGGATGGCTC AAATGCAACC ATGATGCAAG
2651 CTTTGTGCGAT TGACCCAAAG TGAGAGAACT AGATCGAGAT GGATTATTCTG

Fig. 15-2
SUBSTITUTE SHEET (RULE 26)

23/29

SEQ ID NO: 3

2701 GAACCATTAC CGCACCCCTTA TATAATGGCA GCATCTTAAT AGTAAACAAA
2751 AGCTTTAGCC TTAGGTTTTA GCTTCCTTCA CTCTTTGCAT ACATTGTGAA
2801 TCTGCGGTTT TAGATGGACC ATAGTGGAAA AAGGCTTTCA TCAATAACTC
2851 GTGGACTTGA TCAATGGTAG AAAAGANAAT ACATAGTATG GAAAACTAGA
2901 TATTTGATAT ATTTGGTTCA AACTCTTATC CGGTGTTGAG GTGATATACA
2951 CATGAAGACA TAACAATCGC ATAGCCGAGA AACTAGTATT CATTAACCTT
3001 TTTCTCTAAA GAGATTGTCC TATCAATCTA AATTTTAGAT GTTAAAAAAA
3051 AAATggttaag gttaaacagg ccgctagggtt ggtttttacga tgatgtaaaa
3101 agtagccatc ttaaaataac agtcgtttgc gagactggcc aggccatccc
3151 atggggccata ggctcgctca agttgTGCTT GGCAGAATTT AGTAACTTGG
3201 GGTTTTGTTA TCAACAATCA ATAGTTTAAG GCTTTACCTG CAAGAAATGA
3251 AGAGTTTAAG GGTTCCTTTT GGTATTCCCG ATTCACACAA GTGAGCTAGC
3301 TCATCAGAGT CCACGAGCTT CCCACTAAAA AATTGAAAAT TGTTGCTTCT
3351 GTCATCTGAA ^{ELL} ATTAACAAA GCGAGAAAAG GCGATACAAA CGATTTCGAA
3401 TGCTTCATCT TCTCCTTTGA AAATCCTTCT TCTGCTTAAT GCTGCTAGAT
3451 ATGGATCTCG GTGTTCTTCT TCCATCATTG CAATCTGTGA GCTGTCTCTT
3501 TAGCTTTTGA CTGTTGCAAT TGTTATTGTG AAATTTTTGT TCGCTTTTGG
3551 ATCAGCTTTT GTTAAATTCG TTCCGAGATT TTAGGTTTAT GTGCTGGTGT
3601 TTTACTTCGT TTACTTGGGN CGNTGGCGGA GAAATTCCTC CCGGGAAAGT
3651 TATTCGCGGC GTCCTTTTAT CAGATGGCTC TCAACTTCGT TACCGATGCA
3701 ATGGTATATT TGATTTGATT TACTCTCTCT ACAATTCCTG AGAGTCTGTG
3751 AGCTCGAAAG TTCATTTCCA TTAGTTTGGT TAATTCAATT TCAGGTCTAT
3801 TGGCACTAAT ATTGTTGGTA GCTATTTNGG GAATCTGTGC AAAACTTGGC
3851 ATTGTATCAC CTCTTGTAAG TGTAGTTACA AGATTTCGAT TGTATTTCTA
3901 TGAATCCGAA TGCTATATGC TATATGAATC CGATTGCAAT TGCTTTCTCA
3951 CACTCATTC CACTGAGATGT TTGGTAGGTG GTTGCGGATA GAGGACTTGA
4001 GTTACTCTCA GCTACTTNNA TTTCTTGTGT TTGGGGAAGA TGATCAATCC

Fig. 15-3
SUBSTITUTE SHEET (RULE 28)

24/29

SEQ ID NO: 3

4051 TTAGTCCGGN GTCTTG GATT TTAGNTGNGT TACCATCAGA TTNGCTTTGG
4101 GTGGTGTGAT TTGTAATCTC CATGATATCT CTTAATATTC TCAGGTGACA
4151 TTAGCATTGT ATGTTACTGG GCGAAGTTCC TCGAATAAGG GTTCTTCCCT
4201 AAAGCCTCAT GTCTCAGGAA ATCTTGTACA TGA CTGGTAC TAACATAATA
4251 CAATTGTAGA TCTGATACTT TCTTGTTACA CAAAATGTTG TTAAAAGTTA
4301 TATATTTTGA CTCCTGCAAG AGCAAAACTA AGAAATAATC TGGTACTATA
4351 TAGAGTTTGA AACACTGAAT TGGACAAGAT GATTCTATAG AACTTCGTAG
4401 AGTGTTGAGT AATTTCTCCT AGAACGGTTG TAGCTTCCTC TTTTTTCCTT
4451 TTAACCGCAG TGACTTTAGC TTTTGGA ACT TTTCTACTGA AACTAGAAGT
4501 TCTGGTTTTG TCTTTC ACTT ATCTCTTCCA AACAACTGCT TCAATTTTTT
4551 CTCATATTGT TTGTTTCATG TGATAGGTGG TTTGGAATAC AGCTGAATCC
4601 TCAGTTTATG AGCATTGATC TCAAGTAATC CATTTTTCTG TTTTTTCTTC
4651 TATTTGTCAG CCAAGGCTAC ATCATTGCTT CAGTTTGTTT CGTACTCAAT
4701 CGAGTGGCAG TTTAATAATG TAATCAGCAG TTATGCATGG TTATGATGAA
4751 TGGGAGTTAT TCCTTG TGTA GGTTTTTCTT TGTCAGAGCC GGGATGATGG
4801 GATGGCTGCT TATCAATCTC TCTATTCTGG CAAAAGTGT GCAGGATGGT
4851 TCCTTGAGTC AGTCGATGAT CTTTACCAGA TCTTCTGTGC GGTAAATTTG
4901 GTTTTTACTT ACAAATCTTG CTTCTTGAAN TCTGATCATC TGTGTTTTGT
4951 TAGTTTTGAT TAGTTTTATA ATTGCAGTTA TATATATTGG ATACTTTGTT
5001 CATGAAGAAT ACATGACCTC TACGTAAGTT CATGGCGTGT TAAGGAAACA
5051 CATTTGTCTT ACCAAAAAAT GACCATTTCG ATTATTACAT CTACTTTGAT
5101 TTTACTCTTT TCAGGTGGGA CATAATTGCA GAGAGACTAG GCTTCATGCT
5151 AGTGTTTGGA GATCTCCTGT GGATTCCTTT CACTTTTAGC ATTCAGGCAT
5201 GTA ACTGTGA GCCTGAACAC AAACAAGATA TTAATTTATC TTATTGACAG
5251 TATCTTCTTG GCATGTTACA GTTATTCTCG GAAACAATAT TGTTC TAGAA
5301 TGCTTGATCA CTCTGTGACT GAATTGTCTT CTCTCTGGTA CAGGGCTGGT
5351 GGCTTTTGCA CAACAAAGTA GAACTAACAA TTCCTGCGAT TG TAGTCAAT

Fig. 15-4
SUBSTITUTE SHEET (RULE 20)

25/29

SEQ ID NO: 3

5401 TGCCTTGTCT TCTTGATAGG GTAAGTTCTG AGACATGGGG TTATTTTCCA
5451 TTCTTACATA TCTACACTAA GAAACCCACT ATTTCTTCTT TGGCAGGTAC
5501 ATGGTTTTTC GAGGAGCTAA CAAACAAAAA CATATCTTTA AGAAGAACCC
5551 AAAAACACCA ATATGGGGCA AGCCTCCAGT GGTAGTTGGT GGAAAGTTAC
5601 TGGTTTCAGG CTATTGGTAT GTTATATTTA TCTTCTCTTG TTTCTTTGCT
5651 TGGTTTCGCC ATCTCTGTGT TTGATTGTTT ATCATGCTGG GAATAAAGAG
5701 TTGAAAGTTC CGCAATGACA CATTTCCGAT AACTTAGGTG CTGTTTTGTA
5751 TATATGACAG GGGAATTGCA AGGCACTGTA ATTACCTTGG CGACTTGATG
5801 CTTGCTCTGT CCTTCAGTTT GCCATGTGGA ATAAGGTACT CCTNCTGCTT
5851 GAGTTCACCT ACAGCTACCA AAATCATGTA GAAACTAATA CCAATATCNA
5901 AACGTTTCGAA GTTGATTTGG CTGACTTAAA GATATTGATC TCTAACCATC
5951 ATTTGAAAAG TCTAAAGCTT TCAAGTTCAT TTCCCAAAGC TGTTTTTATG
6001 ATATTTTCGTC TNGTGTATTC TCAGTTCTCC GGTTCATAT TTCTACCCGA
6051 TATACCTGCT GATACTATTG ATATGGAGAG AACGAAGAGA CGAAGTTCGA
6101 TGTGCAGAGA AGTACNAGGA GATATGGGCA GAGTATCTTA GACTTGTCCT
6151 CTGGAGAATA CTTCCTTATG TTTATTAGAT GTGCCAAGAG CCAATTCATG
6201 AATCCTTTCA GATTCATCCT CTTGTGTCTT ATTTTTCAT TAAATGTGAC
6251 NTGAAATGAT CCCATTATNG CCTNTTATCA ATGCTTGATT GAAACTTTGT
6301 AGTACACGTT TGAGAATTAC TTCAGTCCTT GTTATTATTT TAGCATGGAT
6351 ATCAACATTT TCGGATTTAT TTNTNGGGT ATTTTAAAC CNNAGATTAC
6401 CNAANAAAAC CATTGTTTGA NGTANGATAA TATGGACTTT TTAAGTAAAA
6451 AAAATNCTAN TAGGGGAACA AATNGAAGTT GAATATGGCT GAATNTTTTT
6501 ATGGANAAAA TGGAAACTTT TCCCACTTTG AAATGACAAT NCAAGTTTGG
6551 TGGACNACTT AATCACTGGA AACGTTAATG GCCAACCN

Fig. 15-5
SUBSTITUTE SHEET (RULE 20)

26/29

1 CTGAAATTAAACAAAGCGAGAAAAGGCGATACAAACGATTTAGAATGCTT 50 SEQ ID NO: 11
|||||
3356 CTGAAATTAAACAAAGCGAGAAAAGGCGATACAAACGATTTCTGAATGCTT 3405 SEQ ID NO: 10
|||||
51 CATCTTCTCCTTTGAAAATCCTTCTTCTGCTTAATGCTGCTAGATATGGA 100
|||||
3406 CATCTTCTCCTTTGAAAATCCTTCTTCTGCTTAATGCTGCTAGATATGGA 3455
|||||
101 TCTCGGTGTTCTTCTTCCATCATTGCAATCT..... 131
|||||
3456 TCTCGGTGTTCTTCTTCCATCATTGCAATCTGTGAGCTGTCTCTTTAGCT 3505
|||||
132GTTTATGTGCTGGTGTTTTAC 152 SEQ ID NO: 12
|||||
3556 CTTTTGTAAATTCGTTCCGAGATTTTAGGTTTATGTGCTGGTGTTTTAC 3605
|||||
153 TTCGTTTACTT.GGCCGTTGCCGGAGAAATTCTCCCCGGGAAAGTTATTC 201
|||||:|||||
3606 TTCGTTTACTTGGGNCGNTGGCGGAGAAATTCTCCCCGGGAAAGTTATTC 3655
|||||
202 GCGGCGTCCTTTTATCAGATGGCTCTCAACTTCGTTACCGATGCAAT... 248
|||||
3656 GCGGCGTCCTTTTATCAGATGGCTCTCAACTTCGTTACCGATGCAATGGT 3705
|||||
249GGTCTATTGGCA 260 SEQ ID NO: 13
|||||
3756 GAAAGTTCATTTCATTAGTTTGGTTAATTCAATTTACGGTCTATTGGCA 3805
|||||
261 CTAATATTGTTGGTAGCTATTTTGGGAATCTGTGCAAACTTGGCATTGT 310
|||||:|||||
3806 CTAATATTGTTGGTAGCTATTTNGGGAATCTGTGCAAACTTGGCATTGT 3855
|||||
311 ATCACCTCTT..... 320
|||||
3856 ATCACCTCTTGTAAGTGTAGTTACAAGATTTGATTGTATTTCTATGAAT 3905
|||||
321GTGGTTGCGGATAGAGGACTTGAGTTAC 348 SEQ ID NO: 14
|||||
3956 ATTCCACTGAGATGTTTGGTAGGTGGTTGCGGATAGAGGACTTGAGTTAC 4005
|||||

Fig. 16-1
SUBSTITUTE SHEET (RULE 28)

27/29

349 TCTCAGCTACTTTTATTTTCTGTGTTT..... 375
|||||||:|||||||
4006 TCTCAGCTACTTNNATTTCTTGTGTTTGGGGAAGATGATCAATCCTTAGT 4055
.
.
.
376TGGTGACATTAgC 388 SEQ ID NO: 15
|||||||
4106 GTGATTTGTAATCTCCATGATATCTCTTAATATTCTCAGGTGACATTAGC 4155
.
389 ATTGTATGTTACTGGGCGAAGTtCCTCgAATAAgGtTctTCCCTAAAgC 438
|||||||
4156 ATTGTATGTTACTGGGCGAAGTTCCTCGAATAAGGGTCTTCCCTAAAGC 4205
.
439 CTCATGTCTCagGAAATCTTGTACATGACT..... 468
|||||||
4206 CTCATGTCTCAGGAAATCTTGTACATGACTGGTACTAACATAATACAATT 4255
.
.
.
469GGTGGTTTGGAAATACAGCTGAATCCTCAGT 498 SEQ ID NO: 16
|||||||
4556 ATTGTTTGTTCATGTGATAGGTGGTTTGGAAATACAGCTGAATCCTCAGT 4605
.
499 TTATGAGCATTGATCTCAA..... 517
|||||||
4606 TTATGAGCATTGATCTCAAGTAATCCATTTTTCTGTTTTTCTTCTATTT 4655
.
.
.
518GTTTTTCTTTGTCTCAGAGCCGGGATGATGGGATGG 551 SEQ ID NO: 17
|||||||
4756 GTTATTCCTTGTGTAGGTTTTTCTTTGTCTCAGAGCCGGGATGATGGGATGG 4805
.
552 CTGCTTATCAATCTCTCTATTCTGGCAAAAAGTGTGCAGGATGGTTCCTT 601
|||||||
4806 CTGCTTATCAATCTCTCTATTCTGGCAAAAAGTGTGCAGGATGGTTCCTT 4855
.
602 GAGTCAGTCGATGATTCTTTACCAGATCTTCTGTGC..... 637
|||||||
4856 GAGTCAGTCGATGA.TCTTTACCAGATCTTCTGTGCGGTAAATTGGTTT 4904
.
.
.
638GTTATATATATTGGACTACTTTGTTTCAT 665 SEQ ID NO: 18
|||||||
4955 TTTGATTAGTTTTATAATTGCAGTTATATATATTGGA.TACTTTGTTTCAT 5003
.
.
.

Fig. 16-2
SUBSTITUTE SHEET (RULE 28)

28/29

666 GAAGaAATACATGACCTCTAC..... 685
|||||
5004 GAAGAATACATGACCTCTACGTAAGTTCATGGCGTGTTAAGGAAACACAT 5053
.
.
.
686GTGGGACATAATTGCAGAGAGACTAgGCTTCATGCTAGT 724 SEQ ID NO: 19
|||||
5104 ACTCTTTTCAGGTGGGACATAATTGCAGAGAGACTAGGCTTCATGCTAGT 5153
.
.
725 GTTTGGAGATCTCCTGTGGATTCCCTTTCACTTTTCAGCATT..... 764
|||||
5154 GTTTGGAGATCTCCTGTGGATTCCCTTTCACTTTTCAGCATTTCAGGCATGTA 5203
.
.
.
765CAGGGCTGGTGGC 777 SEQ ID NO: 20
|||||
5304 TTGATCACTCTGTGACTGAATTGTCTTCTCTCTGGTACAGGGCTGGTGGC 5353
.
.
778 TTTTGCACAACAAAGTAGAACTAACAGTTCCTGCGATTGTAGTCAATTGC 827
|||||
5354 TTTTGCACAACAAAGTAGAACTAACAAATTCCTGCGATTGTAGTCAATTGC 5403
.
.
828 CTTGTCTTCTTGATAG..... 843
|||||
5404 CTTGTCTTCTTGATAGGGTAAGTTCCTGAGACATGGGGTTATTTTCCATTC 5453
.
.
844GGTACATG 851 SEQ ID NO: 21
|||||
5454 TTACATATCTACACTAAGAAACCCACTGTTTCTTCTTTGGCAGGTACATG 5503
.
.
852 GTTTTTTCGAGGAGCTAACAAACAAAAACATATCTTTAAGAAGAACCCAAA 901
|||||
5504 GTTTTTTCGAGGAGCTAACAAACAAAAACATATCTTTAAGAAGAACCCAAA 5553
.
.
902 AACACCAATATGGGGCAAGCCTCCAGTGGTAGTTGGTGGAAAGTTACTGG 951
|||||
5554 AACACCAATATGGGGCAAGCCTCCAGTGGTACTTGGTGGAAAGTTACTGG 5603
.
.
952 TTTTCAGGCTATT..... 963
|||||
5604 TTTTCAGGCTATTGGTATGTTATATTTATCTTCTCTGTTTCTTTGCTTGG 5653
.
.
.
964GGGGAATTGCAAGGCACTGTAATTACCTTGGCGACTTGATGCTT 1007 SEQ ID NO: 22
|||||
5754 ATGACAGGGGAATTGCAAGGCACTGTAATTACCTTGGCGACTTGATGCTT 5803

Fig. 16-3
SUBSTITUTE SHEET (RULE 20)

29/29

1008 GCTCTGTCCTTCAGTTTGCCATGTGGAATA..... 1037
|||||
5804 GCTCTGTCCTTCAGTTTGCCATGTGGAATAAGGTACTCCTNCTGCTTGAG 5853
.
.
.
1038AGTTCTCCGGTTCCATATTTCTACCCGATAT 1068 SEQ ID NO: 23
|||||
6004 TTTCGTCNGTGTATTCTCAGTTCTCCGGTTCCATATTTCTACCCGATAT 6053
.
1069 ACCTTCTGATACTATTGATATGGAGAGAACGAAGAGACGAGGTTTCGATGT 1118
|||||
6054 ACCTGCTGATACTATTGATATGGAGAGAACGAAGAGACGAAGTTTCGATGT 6103
.
1119 GCAGAGAAGTACAAGGAGATATGGGCAGAGTATCTTAGACTTGTCCTCCTG 1168
|||||:|||||
6104 GCAGAGAAGTACNAGGAGATATGGGCAGAGTATCTTAGACTTGTCCTCCTG 6153
.
1169 GAGAATACTTCCTTATGTTTATTAGATGTGCCAAGAGCCAAGTCATGAAT 1218
|||||
6154 GAGAATACTTCCTTATGTTTATTAGATGTGCCAAGAGCCAATTCATGAAT 6203
.
1219 CCTTTCAGATTCACCTCTTGTTGTCTTATTTTTTCCATAA..... 1258
|||||
6204 CCTTTCAGATTCATCCTCTTGTTGTCTTATTTTTTTCATTAAATGTGACNTG 6253
.
1259TCTTGTTTATTTTAGCAATGCTCGAATTGAACTTTGTAG 1299 SEQ ID NO: 24
||:|:|||||
6254 AAATGATCCCATATNGCCTNTTATCAATGCTTG.ATTGAACTTTGTAG 6302
.
1300 TACACTTTTGAAAAATAACTTCAGTCCTT 1328
|||||
6303 TACACGTTTG.AGAATTACTTCAGTCCTT 6330

Fig. 16-4
SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/10644

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12N 9/02, 15/53, 15/63, 1/21, 15/09, 15/10; C12Q 1/68, 1/26

US CL : 526/23.2, 23.6; 435/189, 252.3, 320.1, 419, 6, 91.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 526/23.2, 23.6; 435/189, 252.3, 320.1, 419, 6, 91.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, SCISEARCH, LIFESCI, BIOTECHDS, BIOSIS, EMBASE, CAPLUS, NTIS, WPI, CABA, AGRICOLA
search terms: sterol reductase#

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	TATON M. Microsomal $\Delta^{8,14}$ -Sterol Δ^{14} -Reductase in Higher Plants. Eur. J. Biochem. November 1989. Vol. 185. pages 605-614. see entire document.	1-18, 24, 25
Y	GOAD L.J. Application of Sterol Synthesis Inhibitors to Investigate the Sterol Requirements of Protozoa and Plants. Biochem Soc. Trans. 1990. Vol. 18 pages 63-65. see entire document.	1-18, 24, 25
A	LAI, M.H. The Identification of a Gene family in the Saccharomyces cerevisiae Ergosterol Biosynthesis Pathway. Gene 1994. Vol. 140. pages 41-49.	1-18, 24, 25

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

20 AUGUST 1997

Date of mailing of the international search report

04 SEP 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Authorized officer

REBECCA PROUTY

Facsimile No. (703) 305-3230

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/10644

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-18, 24 and 25

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/10644

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-18, 24 and 25, drawn to C-14 sterol reductase, DNA therefore, and methods of producing C-14 sterol reductase.

Group II, claim 19, drawn to C-14 sterol reductase antibodies.

Group III, claims 20-23, 30 and 31, drawn to transgenic plants.

Group IV, claims 26-29, drawn to a method for reducing C-14 sterol reductase expression.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The inventions of Groups I-III comprise different products with unrelated chemical structures. The method of group IV lacks a corresponding special technical feature because it neither makes nor uses the products of any of Groups I-III.

FIGURE 1A

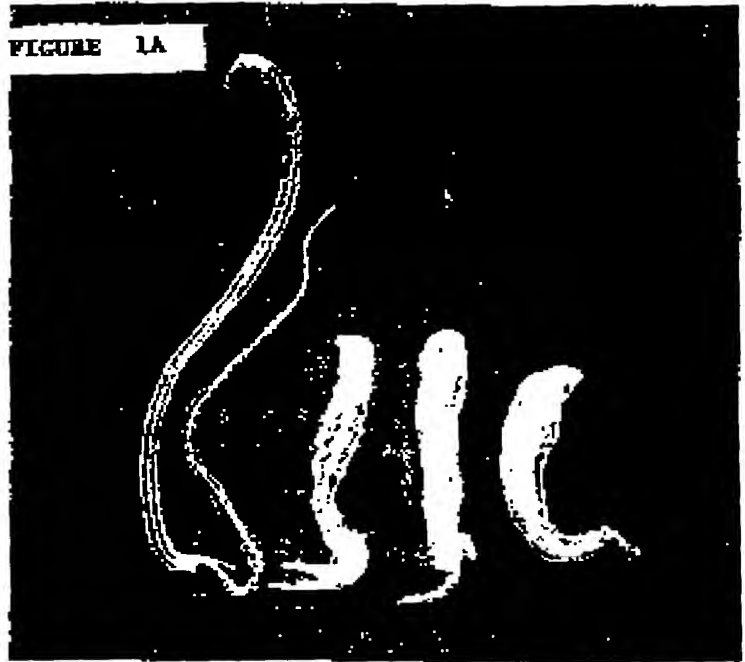
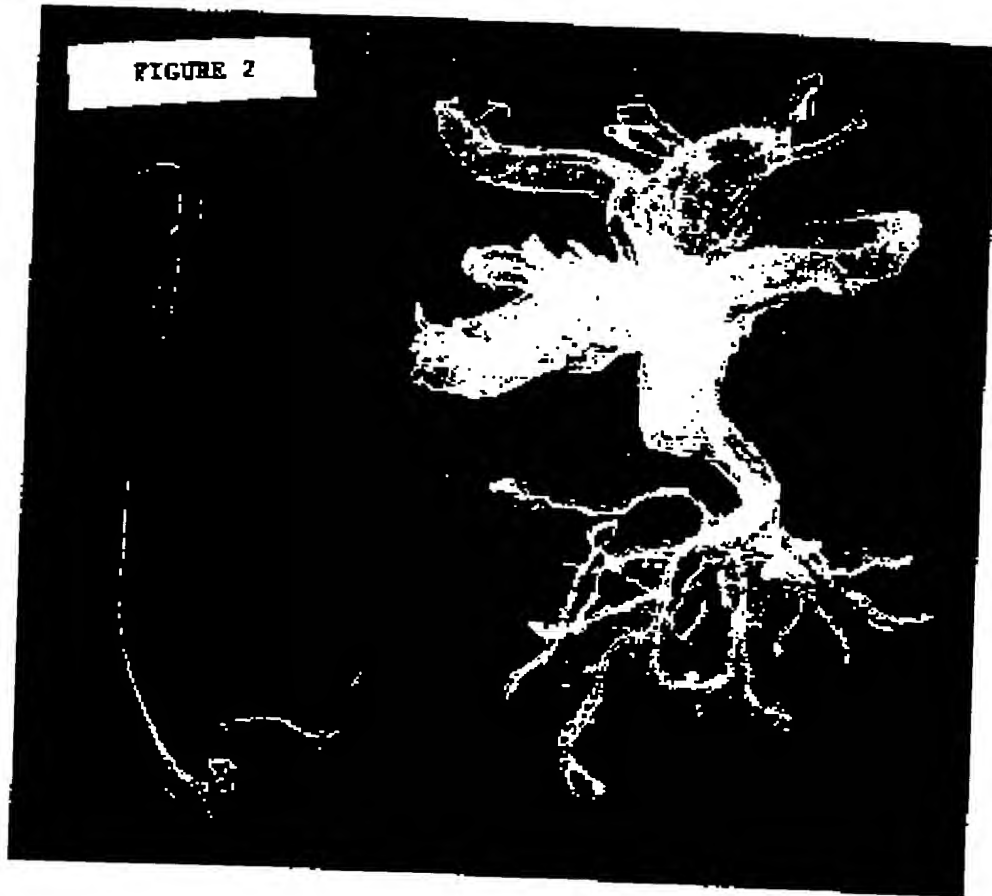
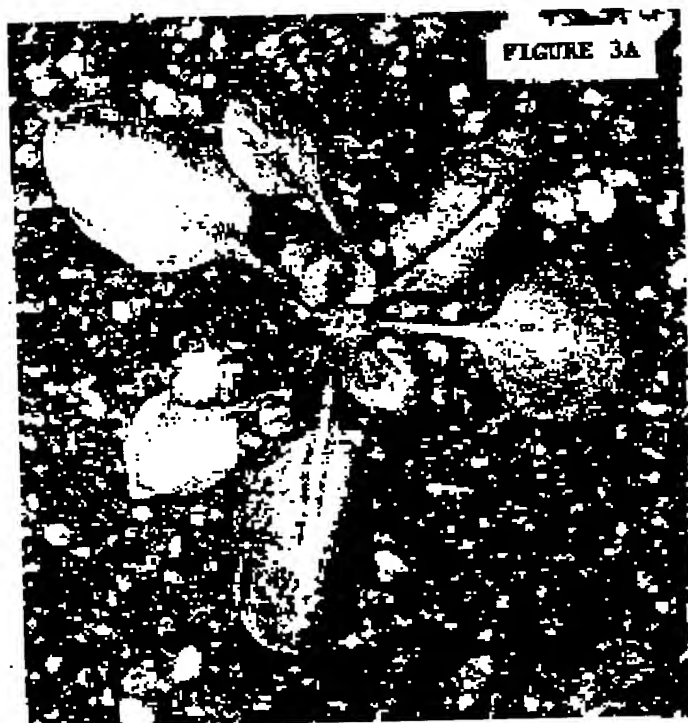


FIGURE 1B



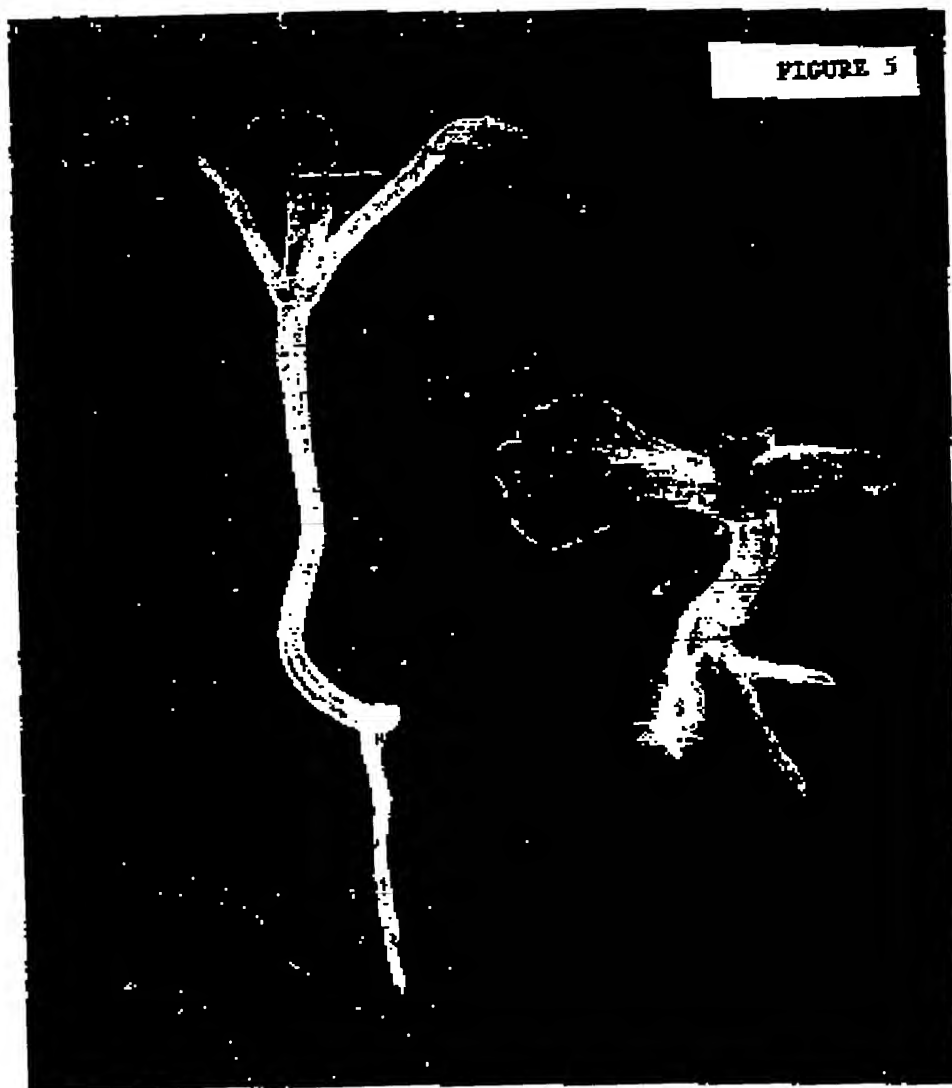




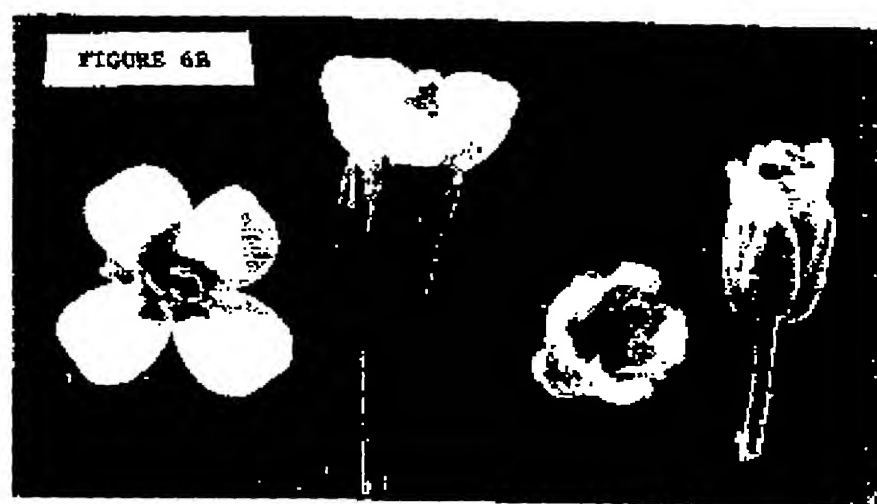
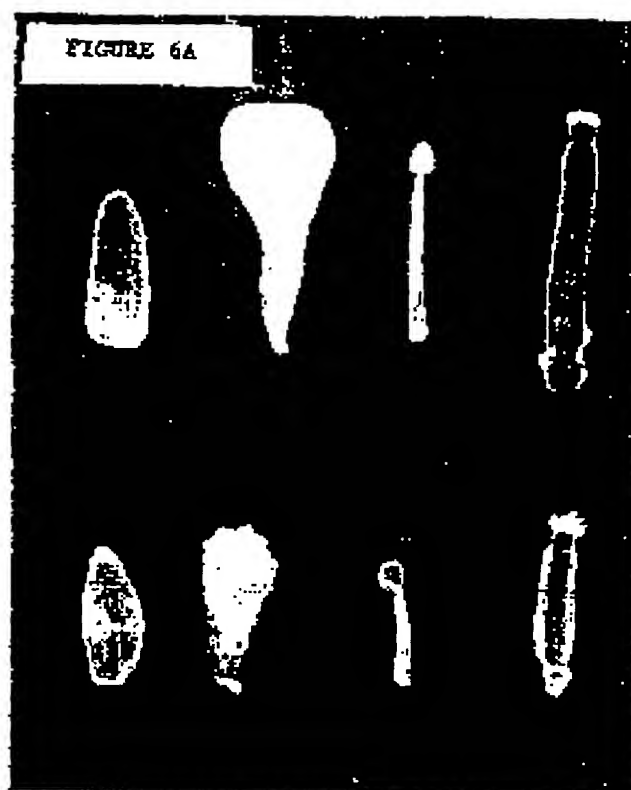
4/29

FIGURE 4





6/29



7/29

FIGURE 7A



FIGURE 7D



FIGURE 7B

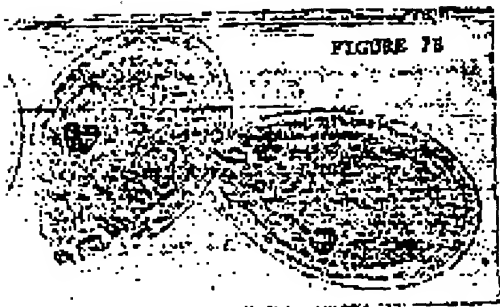


FIGURE 7E



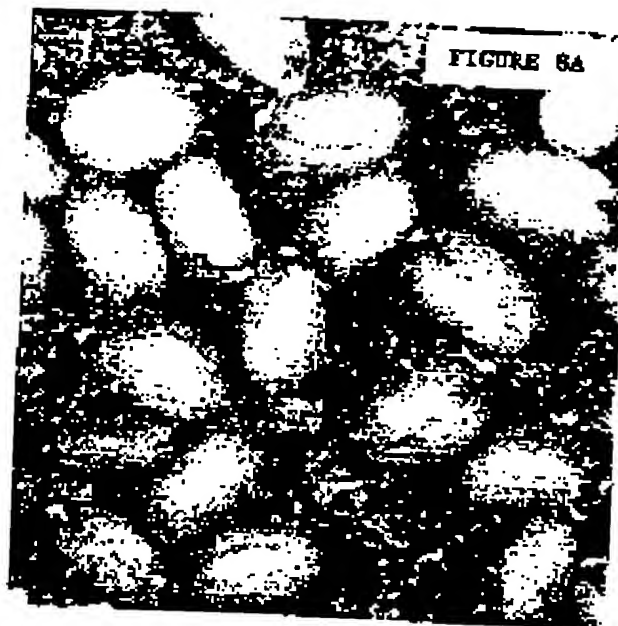
FIGURE 7C



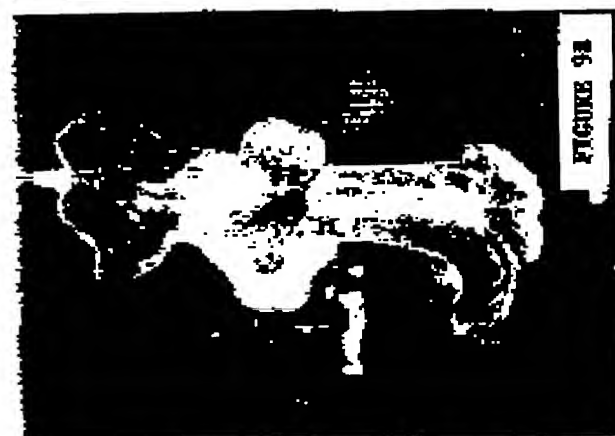
FIGURE 7F



8/29

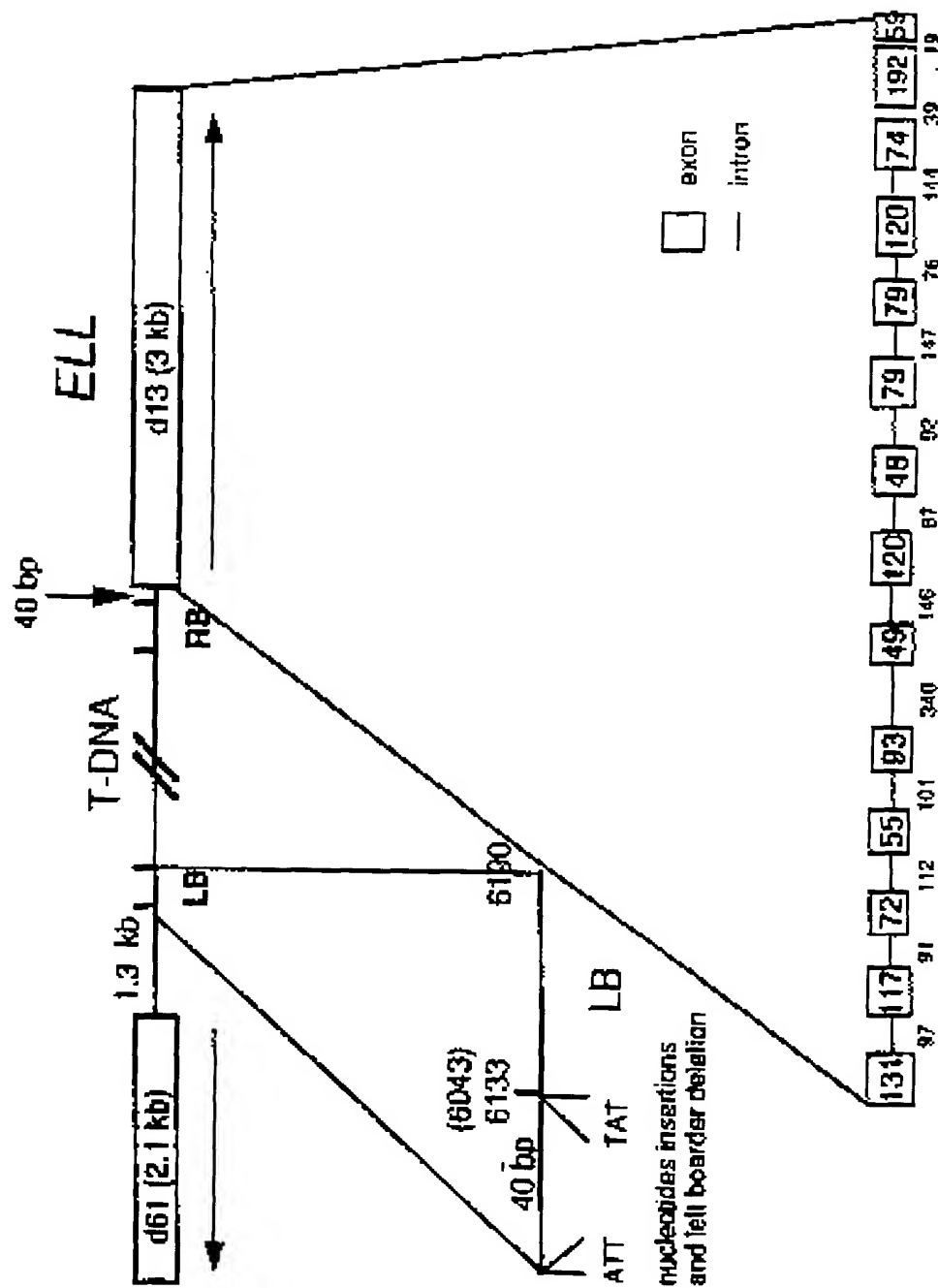


9/29



10/29

ELL encodes a putative C-14 sterol reductase



deduced peptide sequence predicts
8-9 putative transmembrane domains

FIGURE 10A

11/29

chr3

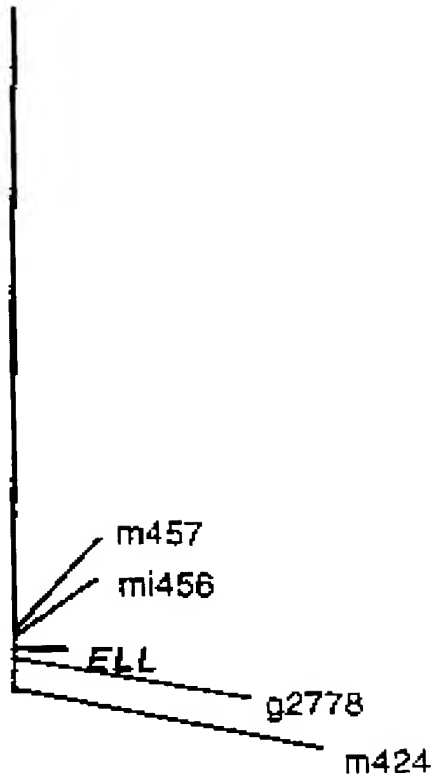


FIGURE 10B



14/29

FIGURE 13A

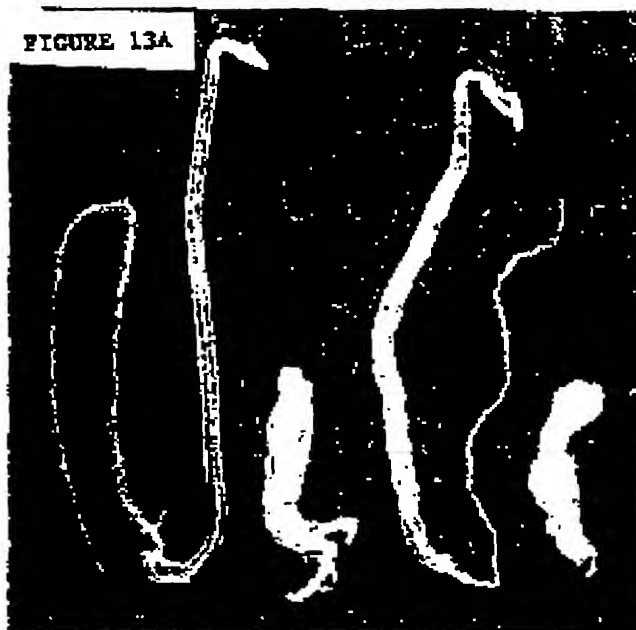


FIGURE 13B

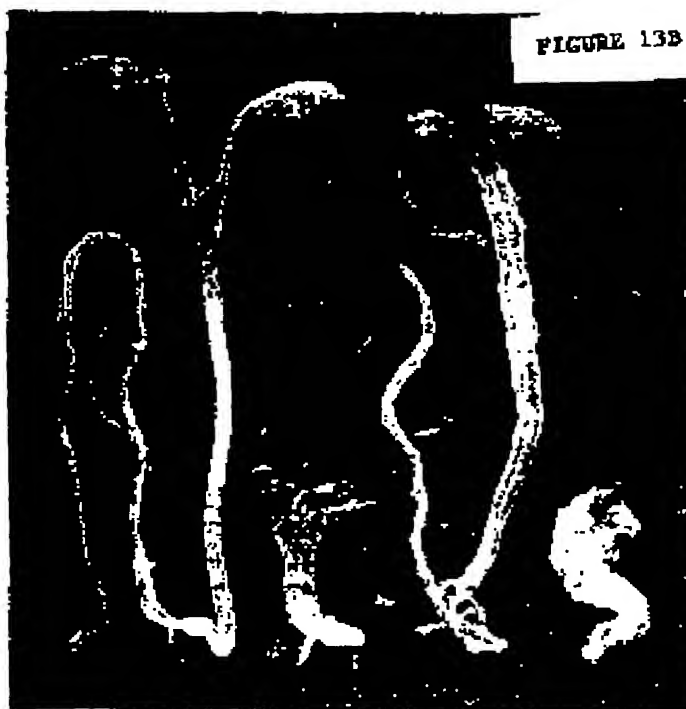


FIGURE 14 / SHEET 1

5

BNSDOCID: <WO_9748793A1TI_>

16/29

FIGURE 14 / SHEET 2

191 ----- 240
AAGAGCGCGCCCTTCAATAAGCGCGGCAGGAAAATAGTCTACCGAAGCTTGAAGCAATCG
c L P G K V I R G V L L S D C S Q L R Y R -
C B C H C
v s S AV iT v
i r s ii nE i
R D P uJ ci R
I I I II II I
/ /
GATGCAATGGTCTATTGGCACTAATATTGTTGGTAGCTATTTTGGGAATCTCTGCAAAAC
241 ----- 300
CTACGTTACGAGATAACCGTGATTATAACAACCATCGATAAAACCGTTAGACACGTTTTC
c C N G L L A L I L L V A I L G I C A K L -
H M B
g a H C c
CH i M AM C e Bi D Av e
jP E n cn j I pn d li s
eh I l il e I l4 e uJ 3
II I I II I II I
/ / / /
TTGGCATTGTATCACCTCTTTGTGGTTCCCGATAGAGGACTTTGAATTACTCTCAGCTACTT
301 ----- 360
AACCGTAACATAGTGGAGAACACCAACGCGCTATCTCTGAACTCAATGAGAGTGGATGAA
c G I V S P L V V A D R G L E L L S A T F -
MT M
as a
ap H e B B T b
l4 p I f s a o
l5 h I i r g I
II I I I I I I
/ / / /
TTATTTTCTGTGTTTTGGTGACATTAGCATTGCTATGTTACTGGCGGAAAGTTCCTCGAATA
361 ----- 420
AATAAAGACACAAAACCACTGTAATCGTAACATACAATGACCGCGTTCAAGGAGCTTAT
c I F C V L V T L A L Y V T G R S S S N R -
N A N
C i l B l
M v a D MW s RT a B
n i I d n2 r sa I s
l J I e l6 G at I r
I I I I II I I
/ /
AgggGtTctTCCCTAAAGCCTCATGTCTCAAGAAATCTTGTACATCACTGGTGGTTTGGAA
421 ----- 480
TCCGCAAGaAGCGATTTCGGAATACAGAGTCTTTTGAACATGTACTGACCACCAACCTT
c C S S L K P H V S G N L V H D W W F G I -
M B S
CsRP c Ha
AvpiVT D e M l u P C S
liAnuf d B n n 3 p vMNCB
uJlflil e 3 l 4 A n lscrB
jllllll i I I I I I JpiPv
llllll i I I I I I lllll

17/29

FIGURE 14 / SHEET 3

// // /
 TACAGCTGAATGCTCACTTTATGAGCATTGATCTCAAGTTTTCTTTGTCAGAGCCGGGA
 481 ----- 540
 ATGTCGACTTAGGAGCTCAAATACTCGTAAGTCTAGAGTTCAAAAAGAAACAGTCTCGCCCT

c Q L N P Q F M S I D L K F F F V R A C M -

			F						
			Cn					C	DN
B	B	vufT	F					v	B rl
c	c	lqos	o					i	c da
c	c	JHka	k					R	c II
I	I	IIII	I					I	I IV

// /
 TGATCGGATGGCTCGTTATCAATCTCTCTATTCTCGCAAAAAGTCTGCAGCATGCTTCCT
 541 ----- 600
 ACTACCGTACCCACGAATAGTTAGAGAGATAAGACCGTTTTTCAGACGTCTTACCAAGGA

c M G W L L I N L S I L A K S V Q D C S L -

			B	S
K		H	CM	BB a
iF	B	TP	eb	g+Du
no	s	al	so	ltp3
fk	q	q*	3I	IYnA
II	I	II	II	IIII

/ / /
 TGAATCACTGCATGATTCTTTACAGATCTTCTGTCGGTTATATATATTGGACTACTTTG
 601 ----- 660
 ACTCACTCAGCTACTAAGAAATGGTCTAGAAGACACCCAATATATATAACCTGATGAAAC

c S Q S M I L Y Q I F C A L Y I L D Y F V -

				T				
			N	a	A		t	N
	l	IM	B	p	1C	B	b	l
R	a	ab	BBT M	5	wv	s	B	ab
c	I	Io	aaa n	0	2i	m	f	if
a	I	II	AlI l	9	6R	F	a	Ia
I	I	II	III I	I	II	I	I	II

/
 TTEATGAAQAATACATGACCTCTACCTGGGACATAATTGCAGAGAGACTAGGCTTCATGC
 661 ----- 720
 AAGTACTTCTTATGTACTCGAGATGCACCTCTATTAAAGTCTCTCTCGATCCGAAGTACG

c H E E Y K T S T W D I I A E R L G F M L -

			S					
C	BB a	H		C		C	C	
M	g+Du	IT		j	B	M	v	v
o	ltp3	nf		e	s	w	i	i
l	IYnA	fi		P	m	o	J	J
I	IIII	II		I	I	I	I	I

/ / /
 TAGTGTITGAGATCTCTCTGTGGATTTCCTTTCACTTTTAGCATTCAGGCTCTGCTGCTTT
 721 ----- 780
 ATCAGAAACCTCTAGAGGACACCTAAGGAAACTGAAARTCGTAAGTCCCGACCACCGAAA

c V F G D L L W I P Y T F S I Q C W W L L -

T

FIGURE 14 / SHEET 4

C		T		S	
V		S	A	P	M
I		P	I	M5	Bb
R		4	w	u0	bo
I		C	N	n9	sl
		I	I	II	II
				/	/

781 TGCACAACAAAGTAGAACTAACACTTCCTCCGATTGTAGTCAATTGCCCTTGTCTCTTGA 840

ACGTCTTGTTCATCTTGATTGTCAAGGACCCCTAACATCAGTTAACCGAACAGAACT

C H N K V E L T V P A I V V N C L V F L I -

					T
					t
					h
					l
R	AM	T	Av	*	IM
s	In	a	li	e	ls
a	Il	q	uJ	R	le
I	II	I	I2	I	II
					/

841 TAGGGTACATGGTTTTTCGAGGAGCTAACAAACAAAACATATCTTTAAGAAGAACCCAA 900

ATCCCATGTACCAAAAAGCTCCTCGATTGTTTTGTTTTGTATAGAAATCTTCTTGGGT

C Q Y M V F R G A N R Q K H I F K K N P K -

						M		
M				CC		CT		C
b	EM	B	C	av	B	JMs	*	C
o	ps	*	j	ci	s	enp	I	j
I	ml	l	e	HJ	r	P1R	I	e
I	II	I	I	II	I	III	I	I
						/		/

901 AAACACCAATATGGGCAAGCCCTCCAGTGGTAGTTGGTCCAAAGTTACTCGTTTCAGCCT 960

TTTGTGGTTATACCCCGTTCCGGAGGTCACCATCAACCACCTTTCAATGACCAAACTCCGA

C T P I W G K P P V V V G C K L L V S C Y -

						T		
						t		
						h		
						E		
C	P	C		s	PT	B	S	l
j	5	v		p	5s	ssf	1	
e	0	i		4	Op	sta	1	
P	9	R		C	9R	JyN	I	
I	I	I		I	II	III	I	
						/		/

961 ATTGCGCAATTGCAAGCCACTGTATTACTTTGGCGACTTGATGCTTGTCTGTCTTGA 1020

TAACCCCTTAACGTTCCGTGACATTAAATGGAACCCGCTCAACTACCAACGAGACAGAACT

C W G I A R H C N Y L G D L M L A L S F S -

						N		
						l		
						a		
						I		
						I		
						I		
						/		/

N

l

a

I

I

I

X

m

n

I

B

sM

as

Wp

II

N

l

a

I

v

FIGURE 14 / SHEET 5

GTTTCGCATGTGGGAATAAGTTCTCCGGTTCCATATTTCTACCCGATATACCTTCTGATAC
 1021 -----+----- 1080
 CAAACGGTACACCTTATTCAAGAGGCCAAGGTATAAAGATGGCCTATATGGGAAGACTATG

 L P C G I S S P V P Y F Y P I Y L L I L -

 A
 1B
 E Ws M M bT C V RT B
 a 2m n oa i sa s
 r 6B l Ig R at q
 I II I II I II I
 / /
 TATTGATATCGGAGAGAACGAAGAGACGAAGTTGATGTGCAGAGAGGTACAAGGAGATAT
 1081 -----+----- 1140
 ATAACTATACCTCTCTTGGCTTCTCTGCTCCAAAGCTACACGTCTCTTCATGTTCTCTATA

 L I W R E R R D E V R C A E K Y K E I W -

 E
 B EcS
 s D soc B
 m d aRr P
 F e JIF m
 I I III I
 / /
 GGCCAGACTATCTTAGACTTGTCCCTCGGAGAATACTTCTTATGTTTATTAGATGTGCC
 1141 -----+----- 1200
 CCCGTCTCATAGAATCTGAACAGGGGACCTCTTATGAAGCAATACAATAATCTACACGG

 A E Y L R L V P W R I L P Y V Y *

 N
 C H1 M M
 v R iaT.X N iT M
 i c nIf m p n n
 J e III n h Ei i
 I I III I I II I
 // /
 AAGAGCCCAAGTCATGAATCCCTTTCAGATTCACCTCTTCTTCTTATTTTTCATATC
 1201 -----+----- 1260
 TTCTCGGTTCACTACTTACGAAAOTCTAAGTGGAGAACACAGAAATAAAAAAGGTATTAG

 T
 s E
 B p c
 sT5 RT o
 t40 sa 5
 Dq9 at 7
 III II I
 / /
 TTGTTTATTTTACCAATGCTGGAATTGAACTTTCTAGTACACTTTTGAAAAATAACTT
 1261 -----+----- 1320
 AACAAATAAAAATCGTTACGAGCTTAACFTTGAACATCATGTGAAAACTTTTATFTGAA

FF KM
 C BnnGas
 BAEvA*uud*pNTCT

FIGURE 14 / SHEET 6

5

2

AcII	AluI	AlwI	Alw26I	AlwNT	ApoI	AvaI	EbsI
BbvI	BcoI	Bce83I	BcoFI	BfaI	BfiI	BglII	BplI
BpmI	BsaAI	BsaBI	BsaKI	BsaJI	BsaWI	BseRI	BsgI
BeIEI	Bali	BamI	BamBI	BamFI	BsrI	BsrDI	BsrQI
SatVI	Cac8I	CjeI	CjePI	CviJI	CvIRI	OdeI	DpnI
OrdII	BaeI	BaqI	EaeI	Eco57I	EcoRII	Fau4KI	FokI
GdiII	HaeIII	HgaI	Hgi2II	Hin4I	HinfI	HphI	MaeIII
MboII	MnlI	MseI	NelI	HspI	HspAII	MunI	NwoI
NciI	NlaIII	NlaIV	NotI	NspV	PleI	PvuII	RcaI
RsaI	Sau3AI	SbfI	SfaNI	SmaI	SspI	StyI	TaiI
TaqI	TatI	TauI	TfiI	ThaI	TseI	Tsp4SI	Tsp4CI
Tsp509I	TspRI	Tth1111I	XbaI				

AatII	AacI	AceIII	AflIII	AflIII	AhdI	ApaI	ApaBI
ApalI	AacI	AvaiI	AvrII	BaeI	BamHI	BanI	BanII
BcgI	BcgI	BclI	BglI	BmgI	Bpu1DI	Bpu1I02I	BsaI
BsaXI	BsaI	BacGI	BstHKA1	Bsp24I	Bsp24I	Bsp1286I	BspEI
BspGI	BspLU11I	BapMI	BsrBI	BstFI	BaeKII	BaeSI	Bst1107I
BstEII	BstXI	Bau36I	ClaI	DraI	DraIII	OrdI	DsaI
EcoI	Eco47III	EcoNI	EcoO109I	EcoRI	EcoRV	FauI	FseI
FspI	HaeI	HaeII	HbaI	HincII	HindIII	KpaI	KpnI
MluI	KmeI	MscI	NarI	NcoI	NdeI	NgoAIV	NheI
NruI	HaeI	NspI	PacI	Pfl1108I	Pfl1MI	PinAI	PmeI
PmlI	PpuMI	PshAI	Psp1406I	PstI	PvuI	RleAI	RsrII
SacII	SalI	SanDI	SapI	Sau96I	ScaI	SmaAI	SfcI
SfiI	SgfI	SgrAI	SimI	SnaBI	SpeI	SphI	SrfII
See4387I	Sse8647I	SetI	StuI	SunI	SwaI	TaqI	TaqII
Tth111I	UbaDI	VspI	XbaI	XcmI	XhoI		

FIGURE 15 / SHEET 1

E11.Tmp6 Length: 6588 May 3, 1996 10:36 Type: N Check: 7899

```

1  TTTTGAACGN TNAAGAAAA NTANGGTAAG CTCCGNAGGA CAAGANTTCT
51  TGTNACCACA ACACAACAC GCCATGACCC NATCGGTTTC TTNTGTTNG
101 AGATCACCTT TCTTGAGTTG GTGGTTTCTG AGHTCAAAGT CCTTCTTGAC
151 TCAGTGAAGT CCAGATGCAG CNTCAAAACT TTTGTCTGT AGACNTAGCA
201 AGAGTAACAG CACCAACCAA ATCGCTATCC CATCTAATCA AAGCCTTATC
251 ACNTTCATCG TCTCATATA TAATCTGAGG CCGTTGTTCC ACATTGTTAT
301 TGTCCCTGCG AATTCTTTGC ATCACAATAC CCATCAGCTN TTCGAGGTTT
351 TCAGCTCCAG AAGTAACCG ATGTACACGG CCTTAAGGT CTTCAAATTT
401 GAACGAAAAC GATTTCCTTA GTCTAGAGA TGGTAAGAA CTGAGCTTCC
451 CTATATCTGA ATCATGCATC ATTGCCGACA TTTCACTTTC AGTGTGAGAA
501 TCATCAGCTG GCTCTAAGCC AAGAGCTGAA TCCCAAAATT TCTGCATCAT
551 CGTCTTTGCC ATATCATTTA CAGCTCCAGA ACTGTTCTCC ACCATTGAAA
601 TAGCTGCGTG AGTAATCTGA AGAAGCTCTA CAGAGCTCC AGCTGATCCA
651 TCTTATCTA TAATTGGAAG ATGTAGAAAC TTTCATCAT GCATTCTATC
701 CAATGCATCC AGAATCGTTG TCTCTAGCGA TGCACATTCA GCAATCCGTC
751 TCATTACCTT CTGACAAGA GTCAATTGAG CAGATAAATT TTGTGCCACC
801 ACTGCGATCA GAATGTCCTT TGAAGTCAAG ATTCCACTCA TTTTGTTCGG
851 CCGTGGAAT GATTAGACAG TTAACCCGCA AATCCCTCAT CCTTTTCGCA
901 GCAACTGAAA CAGGATCTGA TCGTCTACA AGTGCAGCT TCGATCTCTG
951 TAATAATCGT GACAAGCCGG GTTAACATT NTNTCCTTCA AGGTTTCAAT
1001 GAAAGCATAC GTTGCAAGT ATCGGCTTCC CCAATTGTTN TCCACACCTT
1051 CCACTGCAGC AGCTAAAGCA CTACCTTGCT CTGCAGTTTC TCCATCCTAG
1101 AATAGCATC ATACAAACAC TTCTATATAT CCAACAAAGC AATGACTTCA
    
```

FIGURE 15 / SHEET 2

1151 CCATCTCTCCA CAACAGCCAA GTCTCTAAAC TTCCCTTGAA CCATCTTCTC
 1201 AAGAGCCTCA AGCGCCAACG AATCAGAAGT AACAAAAATA GGATTCCCTAG
 1251 TCATAACCTT AGAGACCCAA GTTTGATCCG GTCTCAACCC TTCACCAATC
 1301 ACTCTTGTAG CTACATCTTT ATCAGTAACA ATCTTGAAAA GAAGCCCACT
 1351 TGAATCACTC AACAAACAAG CATCAACAAG CCTAGCAGCC ATTNTTCGAC
 1401 AAGCATCGAA AACAGTAGTT CCTTCNAGGA ATAGTAAGAG CTTTCGATAA
 1451 CCTAAGCTTC TTCGCTGTCT CTCTCCATTA GAAGGACCTT GAGATTGAGG
 1501 TTGAGGACGA GGTGAATTGG GTTTTCAGGT GTTCCCANTA ACACTTCCAT
 1551 TCTCTGATTC TACTCGTTTC TTAGAAGGTG GTGCTCTCTN CCTACACTA
 1601 GAATTGCTTC TCCTCCTCTA TGTGGAAGAA GGACCCCTCG CTTCAGTACT
 1651 CATATTCGGT CAATCTAGGG TTTACTTAGA TCTTAAATCC CTCANAAATC
 1701 ATTCCCTTAG ATATCAAACT CGTCTCTCCA AATGAAAAAT TCAACCTTTA
 1751 ATTCACAAAAC TATTGAAATT TCATCTAAAG CACGAATCTC AATAAAACCC
 1801 AATTGACAAAT AAGACGATT TGCTCTGAGA ATACCATGCA ACATBCACGA
 1851 AAAGGATTCC AATTTAACGG ACGAGGGAAA TGAAACAACT TGAACCCCTA
 1901 AGGATTTCAG CAGAACTTAT GTGGGAAGAT TGGGNATTTA GCGTTTACCT
 1951 TCTTCTTTCT TCTCAAGCT CTCTCTCTCG AGCACTTTCT TTNCCCCAAA
 2001 AACNAACGGC TCTTAACAAT TGAGTTAANC CANTTATCGA GTTTTCATTG
 2051 GNTGTTCCTG TTTCCCGCTG TGTGGTGGNT CACCACCTCG TTTCTTATAA
 2101 TENACGACTA AAAATGTTAA ANATAANACT AANATTTCTT TCTANAAAAA
 2151 TCGTAAANAC CAAATGTTTT TTTTITCTG ATAATGTCT ATAAATCACC
 2201 CTTCTTTTTT AATAATGAA ATTTGATGAC ATTTATCTCT TGTATCTAGN
 2251 AGAGTTAATG GCTAACATAA ANACCAAAAA AATTAATTC NAATAAATAT
 2301 GATTTGTGTG GGTACATGG AAAATTTCTC AATAAATAAA NCAAAAAAAA
 2351 ATTGTATAGA TGCAGTGCAG GTTGTTTCTC GTCAACTTCT GGTGAGCCT
 2401 CACAACCTGT TGTACAAAGT GGACTCCGAT GTAAATTCCT CTTTTAATAA
 2451 CTTACCAATT ACACCATCCA ACATCTGATT TCAAGAAAAA AATTTTTACT
 2501 GAAATGTGAT CGGTCCACAT TTTTCTATGT ACGTTTARGC CTTTAAGGTA
 2551 GACCTTTAAT CCNAAAAAT, CCTGCAATAA CACACCCGAT TANTGGAAGC
 2601 AAGTAGATAC CTCTCTCGTT TCGATCGCTC AATGCAACCC ATGATCGAAG
 2651 CTTTTCGGAT TGACCCAAAG TGAGAGAACT AATCGACAT GGATTATTCT

FIGURE 15 / SHEET 3

2701 GAACCAATTAC GGCACCCCTTA TATAATGGCA GCATCTTAAT AGTAAACAA
2751 AGCTTTAGCC TTACCTTTTA GCTTCCTTCA CTCCTTGCAT ACATTGTGAA
2801 TCTGCGGTTT TAGATGGACC ATAGTGGAAA AAGGCTTTCA TCAATAACTC
2851 GTGGACTTCA TCAATGGTAG AAAACANAAT ACATAGTATG GAAACTAGA
2901 TATTTGATAT ATTGGGTCA AACTCTTATC CGGTGTTGAG GTGATATACA
2951 CATGAAGACA TAACAATGCC ATAGCCGAGA AACTAGTATT CATTAACTTT
3001 TTTCTCTAAA GAGATTCTCC TATCAATCTA AATTTTAGAT GTTAAAAAA
3051 AAATggcaag gctaaacagg ccgctagglt gglllccaga cgaatgtaaa
3101 agtagccatc ctgaaataac agtcgcttgc gacacgggco aggcacatcc
3151 atggggccata ggctcgctca agttgTcctt GGCAGAATTT AGTAACTTG
3201 GCTTTTCTTA TCAACAATCA ATAGTTTAAG GCTTTACCTG CAAGAAATGA
3251 AGAGTTTAAG GCTTCTTTT GGTATTCCCG ATTCAACAAA GTGAGCTAGC
3301 TCATCAGAGT CCACGAGCTT CCCACTAAA AATTCAAAAT TCTTGCTTCT
3351 GTCATCTGAA ATTAACAAA GCGAGAAAAG CCCATACAAA CCATTTCCGA
3401 TGCTTCTCT TCTCCTTGA AATGCTTCT TCTGCTTAAT GCTGCTAGAT
3451 ATGATCTCTG GTGTTCTTCT TCCATCATTC CAATCTGTGA GCTGCTCTCT
3501 TAGCTTTTGA CTGTTGCAAT GTTTATTGTG AAATTTTGT TCGCTTTTGG
3551 ATCAGCTTTT GTTAAATTCG TTCCGAGATT TTAGCTTTAT CTGCTGCTCT
3601 TTTACTTCTT TTACTIONN CCGTGGCGGA GAAATTCTCC CCGGGAAGT
3651 TATTEGCGCG GTCTTTTAT CAGATCGCTC TCAACTTCTT TCCGATGCA
3701 ATGCTATATT TGATTGATT TACTCTCTCT ACAATTCTCT AGAGTCTCT
3751 AGCTCGAAAG TCAATTCCA TTAATTCTCT TAATTCAATT TCAGCTCTAT
3801 TGGCACTAAT ATTGTTCTA CCTATTTCG GAATCTGTGT AAAACTTGGC
3851 ATTCTATCAC CTCTTGTAAG TGTACTTACA AGATTTGGAT TGTATTTCTA
3901 TGAATCCGAA TGCATATTC TATATGAATC CGATTGCAAT TGCTTTCTCA
3951 CACTCATTCG ACTGAGATCT TTGCTAGGTG GTTGGCGATA GAGGACTTGA
4001 TTTACTCTCA CCTACTTANA TTTCTTGTCT TTGGGGAAGA TCACTCAATC
4051 TTACTCGCGN CTCTTGGATT TTAGTGTGCT TACCTTCAGA TTGCTTTTGG
4101 GTGCTGTGAT TTGTAATCTC CATGATATCT CTTAATATTC TCAGCTGACA
4151 TTAGCAATCT ATCTTACTCG GCGAAGTTCC TCGAATAAGG GTTCTTCTCT
4201 AAAGCTTCTT GTCTCAGGAA ATCTTGTACA TGACTGCTAC TACCTAATA

4251 CAATTCTAGA TCTGATACTT TCTTCTTACA CAAATGTTC TTAAAGTTA
4301 TATATTTTGA CTCTGGAAG ACCAAAACTA AGAAATAATC TGCTACTATA
4351 TAGACTTTCA AACACTGAAT TGGACAAGAT GATTCTATAG AACTTCBTAG
4401 AGTCTTCAGT AATTTCTCCT AGAACGGTTC TAGCTTCCTC TTTTTCCTT
4451 TTAACCGCAG TGACTTTAGC TTTTGGAACT TTTCTACTCA AACTAGAACT
4501 TCTGCTTTTC TCTTTCACCT ATCTCTTCCA AACAACTGCT TCAATTTTIT
4551 CTCATATTGT TTCTTTCATG TCATAGCTGG TTTGGAATAC AGCTGAATCC
4601 TCAGTTTATC AGCATTCATC TCAAGTAATC CATTTTTCTC TTTTTCCTC
4651 TATTTCTCAG CCAAGGCTAC ATCATTCCTT CAGTTTCTTC CGTACTCAAT
4701 CGAGTGGGAG TTTAATAATC TAATCAGCAG TTATGCATGG TTATGATGAA
4751 TGGGAGTTAT TCCTTCTCTA GGTTTTCTT TCTCAGAGCC GGGATGATCG
4801 GATGCTTCTT TATCAATCTC TCTATTCTGG CAAAAAGTCT CCACCATGCT
4851 TCCTTGAGTC AGTCGATCAT CTTTACCAGA TCTTCTCTCC COTAAATTC
4901 CTTTTTACTT ACAAATCTTG CTTCTTGAAN TCTGATCATC TGTGTTTTOT
4951 TAGTTTGGAT TAGTTTTATA ATTGCACTTA TATATATTGG ATACTTTGTT
5001 CATGAAGAAT ACATGACCTC TACGTAACTT CATGCGCTGT TAAGGAACA
5051 CATTGTCTT ACCAAAAAAT GACCATTTCC ATTATTACAT CTACTTTGAT
5102 TTTACTCTTT TCAGCTGGGA CATAATTCCA GAGAGACTAG GCTTCATGCT
5152 AGTCTTTGGA GATCTCCTGT GGATTCCTTT CACTTTTAGC ATTACGGCAT
5201 CTAACCTGCA CCTGGAACAC AAACAAGATA TTAATTTATC TTATTGACAC
5252 TATCTCTCTG GCATGTTACA CTTATTTCTG GAAACAATAT TGTCTAGAA
5301 TCTTGAACA CTCTGTGACT GAATGTCTT CTCTCTGTA CAGCGCTGCT
5351 GCTTTTTCBA CAAACAAATA CAACTAACAA TTCTTCCGAT TCTAGTCAAT
5401 TGCTTTGTCT TCTTGATAGG GTAAGTTCTC AGACATGGGG TTAFTTTCCA
5451 TTCTTACATA TCTACACTAA GAAACCCACT ATTTCTTCTT TGGCAGGTAC
5501 ATGCTTTTTC GAGGAGCTTA CAAACAAAA CATATCTTTA AGAAGAACC
5552 AAAACACCA ATATGGGGCA AGCTTCCACT GGTAGTTGCT GGAAGTTAC
5601 TGCTTTCAGG CTATTGCTAT GTTATATTTA TCTCTCTCTG TTTCTTTGCT
5651 TGCTTTCGCT ATCTCTCTCT TTAGTTCTTC ATCATGCTCG GAATAAGAG
5701 TTGAAAGTTC CGCAATGACA CATTTCGAT AACTTACCTC CTCTTTTCTA
5751 TATATGACAC GCGAATTGCA AGGCACTGTA ATTACCTTGG CGACTTCATC

25/29

FIGURE 15 / SHEET 5

```

5801  CTTCCTCTGT  CCTTCAGTTT  GCCATGTGGA  ATAAGGTACT  CCTNCTGCTT
5851  GAGTTCACCT  ACAGCTACCA  AAATCATGTA  GAAACTAATA  CCAATATCNA
5901  AACGTTGGA  GTTGATTTGG  CTGACTTAAA  GATATTGATC  TCTAACCATC
5951  ATTTGAAAAC  TCTAAAGCTT  TCAAGTTTCA  TTCCCAAGCC  TCTTTTATG
6001  ATATTTGGTC  TNGTGTATTC  TCAGTTCTCC  GGTTCATAT  TTCTACCCGA
6051  TATACCTGCT  GATACFATTC  ATATGGAGAC  AACCAACAGA  CCAACTTCGA
6101  TGTGCAGAGA  AGTACNAGGA  GATATGCCCA  GAGTATCTTA  GACTGTGCCC
6151  CTGGAGAATA  CTTCCCTTATC  TTTATTAGAT  GTGCCAAGAG  CCAATTTCATC
6201  AATCCCTTCA  GATTCACTCT  CTGTGTCTT  ATTTTTCAT  TAAATGTGAC
6251  NTGAAATCAT  CCCATCTATG  CCTTPTATCA  ATCTTGATT  GAACTTTCT
6301  AGTACACCTT  TGAGAATTAC  ELL ← TCTACTGCTT | GTTATTATTT  TAGCATGAT
6351  ATCAACATTT  TCGGATTTAT  TTNTNGGCT  ATTTTAAAC  CNNAGATTAC
6401  CNAANAAAC  CATTGTTTGA  NGTANGATAA  TATGACTTT  TTACTGAAA
6451  AAAATNCTAN  TAGGGGAACA  AATNGAAGTT  GAATATGGCT  GAATNTTTT
6501  ATOGANAAA  TGGAAACTTT  TCCCACTTTG  AAATGACAAT  NCAAGTTTGG
6551  TGCACNACTT  AATCACTOGA  AACGTTAATG  GCGAACCN

```

26/29

FIGURE 16
SHEET 1

```

1 CTGAAATTAAACAAACCGGAGAAAAGGCGATACAAACGATTTCGAATGCTT 50
|||||
3356 CTGAAATTAAACAAACCGGAGAAAAGGCGATACAAACGATTTCGAATGCTT 3405

51 CATCTTCTCCTTTGAAAATCCTTCTCTCTGCTTAATGCTGCTAGATATGGA 100
|||||
3406 CATCTTCTCCTTTGAAAATCCTTCTCTCTGCTTAATGCTGCTAGATATGGA 3455

101 TCTCGGTGTTCTCTCTTCCATCATTCGAATCT..... 131
|||||
3456 TCTCGGTGTTCTCTCTTCCATCATTCGAATCTCTGAGCTGCTCTCTTTAGCT 3505

132 .....GTTTATCTGCTGCTGCTTTTAC 152
|||||
3556 CTTTCTTAAATTGCTTCCGAGATTTTAGGTTTATGCTGCTGCTTTTAC 3605

153 TTCTTTTACTT.GCCCGTTGCCCGGAGAAATTCTCCCGGGAAAGTTATTCT 201
|||||
3606 TTCTTTTACTTGGGNCGN TGCCCGGAGAAATTCTCCCGGGAAAGTTATTCT 3655

202 GCGGCGTCTCTTTTATCAGATGCGCTCTCAACTTCTTACCGATGCCAAT... 248
|||||
3656 GCGGCGTCTCTTTTATCAGATGCGCTCTCAACTTCTTACCGATGCCAATGCT 3705

249 .....GGTCTATTGGCA 260

```

BNSDOCID: <WO 9748793A1TI >

FIGURE 16 / SHEET 3

638 GTTATATATATTTGGACTACTTTGTTTCAT 665
TT
4955 TTTCATTAGTTTTATAATTGCAGTTATATATATTCGA . TACTTTGTTTCAT 5003
666 GAAQAATACATGACCTCTAC 685
TT
5004 GAAQAATACATGACCTCTACGTAAGTTTCATGGCGTCTTAAGCAACACAT 5053
686 GTGGGACATAATTGCAGAGAGACTAGGCTTCATGCTAGT 724
TT
5104 ACTCTTTTTCAGGTGGGACATAATTGCAGAGAGACTAGGCTTCATGCTAGT 5153
725 GTTTCCAGATCTCCTGTGGATTCTTTTCACTTTTACCATT 764
TT
5154 GTTTGGAGATCTCCTGTGCAATTCTTTTCACTTTTAGCAITCAGGCATGTA 5203
765 CAGGGCTGGTGGC 777
TT
5304 TTGATCACTCTGTGACTGAAITTTCTCTTCTCTGTTACAGGGCTGGTGGC 5353
778 TTTTGCACACAAAGTAAGAACTAACAGTTCTTCCGATTGTAGTCAATTGC 827
TT
5354 TTTTGCACACAAAGTAGAACTAACAAITTCCTGCCATTGTAGTCAATTGC 5403
828 CTTGTCTTCTTGATAG 843
TT
5404 CTTGTCTTCTTGATAGCGGTAACTTCTGAGACATGGGTTATTMTCCATTTC 5453
844 GGTACATG 851
TT
5454 TTACATATCTACACTAAGAAACCCACTATTTCCTTCTTTGCCAGGTACATG 5503
852 GTTTTTTCGAGGAGCTAACAACAAAAACATATCTTTAAGAAGAACCCAAA 901
TT
5504 GTTTTTTCGAGGACCTAACAAACAAAAACATATCTTTAAGAAGAACCCAAA 5553
902 AACACCAATATGGGGCAAGCCTCCAGTGGTACTTCCCTGCAAAAGTTACTGG 951
TT
5554 AACACCAATATGGGGCAAGCCTCCAGTGGTACTTGGTGGGAAAGTTACTGG 5603
952 TTTACGGCTATT 963
TT
5604 TTTACGGCTATTGGTATGTTATATTTATCTCTCTCTGTTTCTTTGCTTGG 5653
964 GGGGAATTGCCAAGCCACTGTAATTACCTTGGCGACTTGATGCTT 1007
TT
5754 ATGACAGGGGAATTGCCAAGCCACTGTAATTACCTTGGCGACTTGATGCTT 5803
1008 GCTCTGTCTTCACTTTGCCCATGTGGAATA 1037

29/29

FIGURE 16 / SHEET 4

5804 GCTCTCTCCTTCAGTTTGCCATGTGGAATAAGGTACTCCTNCTCCTTCAG 5853
.....
1036ACCTTCTCCOCTTCCATATTTCTACCCGATAT 1068
1004 TTTCTGTCTGTGTATTCTCAGTTCTCCOCTTCCATATTTCTAGCCGATAT 6053
1069 ACCTTCTGTACTATTGTATATGAGAGAACGAAGAGACGAGGTTCCGATGT 1118
6054 ACCTTCTGTACTATTGTATATGAGAGAACGAAGAGACGAACTTCCGATGT 6103
1119 GCACAGAAAGTACAAAGAGATATGGGCAGAGTATCTTAGACTTGTCCCTTG 1168
6104 GCACAGAAAGTACAAAGAGATATGGGCAGAGTATCTTAGACTTGTCCCTTG 6153
1169 GAGAATACTTCTTATGTTTATTAGATGTGCCAAGAGCCCAAGTCATGAAT 1218
6154 GAGAATACTTCTTATGTTTATTAGATGTGCCAAGAGCCCAATTCATGAAT 6203
1219 CCTTTCAGATTACCTCTTGTGTCTTATTTTTTCCATAA..... 1258
6204 CCTTTCAGATTACCTCTTGTGTCTTATTTTTTCATTAAATGTGACNTG 6253
1259TCTTGTTTTATTTTAGCAATGCTCGAATTGAACTTTGTAG 1299
6254 AAATCATCCCATTATNCCCTNTTATCAATGCTTG.ATTGAAACTTTGTAG 6302
1300 TACACTTTTCAAAAATAAATTTCAGTCCTT 1328
6303 TACACGTTG.AGAATTACTTCAGTCCTT 6330

